PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/285, C12N 1/21, A61K 39/395, C07K 16/12 // (C12N 1/21, C12R 1:21)

A1

(11) International Publication Number: WO 95/13370

(43) International Publication Date: 18 May 1995 (18.05.95)

(21) International Application Number: PCT/CA94/00616

(22) International Filing Date: 7 November 1994 (07.11.94)

(30) Priority Data:
08/148,968 8 November 1993 (08.11.93) US
08/175,116 29 December 1993 (29.12.93) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/175,116 (CIP)
Filed on 29 December 1993 (29.12.93)

(71) Applicant *(for all designated States except US)*: CONNAUGHT LABORATORIES LIMITED [CA/CA]; 1755 Steeles Avenue West, Willowdale, Ontario M2R 3T4 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LOOSMORE, Sheena [CA/CA]; 70 Crawford Rose Drive, Aurora, Ontario L4G 4R4 (CA). HARKNESS, Robin [CA/CA]; Apartment 1706, 640 Sheppard Avenue East, Willowdale, Ontario M2K 1B8 (CA). SCHRYVERS, Anthony [CA/CA]; 39 Edforth Road N.W., Calgary, Alberta T3A 3V8 (CA). CHONG, Pele [CA/CA]; 32 Estoril Street, Richmond Hill, Ontario L4C 0B6 (CA). GRAY-OWEN, Scott [CA/CA]; 2210-142 Silvergrove Drive N.W., Calgary, Alberta T3B 5H4 (CA). YANG, Yan-Ping [CA/CA]; Apartment 1709, 120 Torresdale Avenue, Willowdale, Ontario M2R 3N7 (CA). MURDIN, Andrew [GB/CA]; 146 Rhodes Circle, Newmarket, Ontario L3X 1V2 (CA). KLEIN, Michel [CA/CA]; 16 Munro Boulevard, Willowdale, Ontario M2P 1B9 (CA).

(74) Agent: STEWART, Michael, I.; Sim & McBurney, Suite 701, 330 University Avenue, Toronto, Ontario M5G 1R7 (CA).

(81) Designated States: AU, BR, CA, CN, FI, JP, KR, NO, NZ, RU, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES

(57) Abstract

Purified and isolated nucleic acid is provided which encodes a transferrin receptor protein of a strain of *Haemophilus* or a fragment or an analog of the transferrin receptor protein. The nucleic acid sequence may be used to produce peptides free of contaminants derived from bacteria normally containing the Tbp1 or Tbp2 proteins for purposes of diagnostics and medical treatment. Furthermore, the nucleic acid molecule may be used in the diagnosis of infection. Also provided are recombinant Tbp1 or Tbp2 and methods for purification of the same. Live vectors expressing epitopes of transferrin receptor protein for vaccination are provided.

TITLE OF INVENTION

HAEMOPHILUS TRANSFERRIN RECEPTOR GENES

FIELD OF INVENTION

The present invention is related to the molecular cloning of genes encoding transferrin receptor and in particular to the cloning of transferrin receptor genes from *Haemophilus influenzae*.

REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part of copending United States Patent Application Serial No. 08/175116, filed December 29, 1993, which itself is a continuation-in-part of copending United States Patent Application 08/148,968 filed November 8, 1993.

BACKGROUND OF THE INVENTION

Encapsulated *Haemophilus influenzae* type *b* strains are the major cause of bacterial meningitis and other invasive infections in young children. However, the non-encapsulated or non-typable *H. influenzae* (NTHi) are responsible for a wide range of human diseases including otitis media, epiglottitis, pneumonia, and tracheobronchitis. Vaccines based upon *H. influenzae* type b capsular polysaccharide conjugated to diphtheria toxoid (Berkowitz *et al.*, 1987. Throughout this application, various references are referred to in parenthesis to more fully describe the state of the art to which this invention pertains. Full bibliographic information for each citation is found at the end of the specification, immediately preceding the claims. The disclosures of these references are hereby incorporated by reference into the present disclosure), tetanus toxoid (Classon *et al.*, 1989 and US patent 4,496,538), or *Neisseria meningitidis* outer membrane protein (Black *et al.*, 1991) have been effective in reducing *H. influenzae* type b-induced meningitis, but not NTHi-induced disease (Bluestone, 1982).

regulated (Morton et al., 1993) and a putative *fur*-binding site (Braun and Hantke, 1991) has been identified upstream of *tbp2*. This sequence is found in the promoter region of genes which are negatively regulated by iron, including *N. meningitidis* TfR (Legrain et al., 1993). The promoter is followed by the *tbp2* and *tbp1* genes, an arrangement found in other bacterial TfR operons (Legrain et al, 1993; Wilton et al., 1993). Antibodies which block the access of the transferrin receptor to its iron source may prevent bacterial growth. In addition, antibodies against TfR that are opsonizing or bactericidal may also provide protection by alternative mechanisms. Thus, the transferrin receptor, fragments thereof, its constituent chains, or peptides derived therefrom are vaccine candidates to protect against *H. influenzae* disease. Mice immunized with *N. meningitidis* TfR proteins in Freund's adjuvant were protected from homologous challenge and the anti-TfR antisera were bactericidal and protective in a passive transfer assay (Danve et al., 1993). Pigs immunized with recombinant *A. pleuropneumoniae* Tbp2 were protected against homologous challenge but not heterologous challenge (Rossi-Campos et al., 1992). These data indicate the efficacy of TfR-based vaccines in protection from disease. It would be desirable to provide the sequence of the DNA molecule that encodes transferrin receptor and peptides corresponding to portions of the transferrin receptor and vectors containing such sequences for diagnosis, immunization and the generation of diagnostic and immunological reagents.

Poliovirus is an enterovirus, a genus of the family Picornaviridae. There are three distinct serotypes of the virus, and multiple strains within each serotype. Virulent strains are causative agents of paralytic poliomyelitis. Attenuated strains, which have reduced potential to cause paralytic disease, and inactivated

for expressing the TfR genes by recombinant DNA means for providing, in an economical manner, purified and isolated transferrin receptor subunits, fragments or analogs thereof. The transferrin receptor, subunits or fragments thereof or analogs thereof, as well as nucleic acid molecules encoding the same and vectors containing such nucleic acid molecules, are useful in immunogenic compositions against diseases caused by *Haemophilus*, the diagnosis of infection by *Haemophilus* and as tools for the generation of immunological reagents. Monoclonal antibodies or mono-specific antisera (antibodies) raised against the transferrin receptor protein produced in accordance with aspects of the present invention are useful for the diagnosis of infection by *Haemophilus*, the specific detection of *Haemophilus* (in for example *in vitro* and *in vivo* assays) and for the treatment of diseases caused by *Haemophilus*.

Peptides corresponding to portions of the transferrin receptor or analogs thereof are useful immunogenic compositions against disease caused by *Haemophilus*, the diagnosis of infection by *Haemophilus* and as tools for the generation of immunological reagents. Monoclonal antibodies or antisera raised against these peptides, produced in accordance with aspects of the present invention, are useful for the diagnosis of infection by *Haemophilus*, the specific detection of *Haemophilus* (in, for example, *in vitro* and *in vivo* assays) and for use in passive immunization as a treatment of disease caused by *Haemophilus*.

In accordance with one aspect of the present invention, there is provided a purified and isolated nucleic acid molecule encoding a transferrin receptor protein of a strain of *Haemophilus*, more particularly, a strain of *H. influenzae*, specifically a strain of *H. influenzae* type b, such as *H. influenzae* type b strain DL63, Eagan or MinnA, or a non-typable strain of *H.*

The vector may be one having the characteristics of plasmid DS-712-1-3 having ATCC accession number 75603 or plasmid JB-1042-7-6 having ATCC accession number 75607.

The plasmids may be adapted for expression of the encoded transferrin receptor, fragments or analogs thereof, in a heterologous or homologous host, in either a lipidated or non-lipidated form. Accordingly, a further aspect of the present invention provides an expression vector adapted for transformation of a host comprising a nucleic acid molecule as provided herein and expression means operatively coupled to the nucleic acid molecule for expression by the host of the transferrin receptor protein or the fragment or analog of the transferrin receptor protein. In specific embodiments of this aspect of the invention, the nucleic acid molecule may encode substantially all the transferrin receptor protein, only the Tbp1 protein or only the Tbp2 protein of the *Haemophilus* strain. The expression means may include a nucleic acid portion encoding a leader sequence for secretion from the host of the transferrin receptor protein or the fragment or the analog of the transferrin receptor protein. The expression means also may include a nucleic acid portion encoding a lipidation signal for expression from the host of a lipidated form of the transferrin receptor protein or the fragment or the analog of the transferrin receptor protein. The expression plasmid may have the identifying characteristics of plasmid JB-1468-29, JB-1600-1 or JB-1424-2-8. The host may be selected from, for example, *Escherichia coli*, *Bacillus*, *Haemophilus*, fungi, yeast or baculovirus and Semliki Forest virus expression systems may be used.

In an additional aspect of the invention, there is provided a transformed host containing an expression vector as provided herein. Such host may selected from JB-1476-2-1, JB-1437-4-1 and JB-1607-1-1. The invention

In accordance with another aspect of the invention, an immunogenic composition is provided which comprises at least one active component selected from at least one nucleic acid molecule as provided herein, at least one recombinant protein as provided herein, at least one of the purified and isolated Tbp1 or Tbp2 proteins, as provided herein, at least one synthetic peptide, as provided herein, and a live vector, as provided herein, and a pharmaceutically acceptable carrier therefor or vector therefor. The at least one active component produces an immune response when administered to a host.

The immunogenic compositions provided herein may be formulated as a vaccine for *in vivo* administration to protect against diseases caused by bacterial pathogens that produce transferrin receptors. For such purpose, the compositions may be formulated as a microparticle, capsule or liposome preparation. Alternatively, the compositions may be provided in combination with a targeting molecule for delivery to specific cells of the immune system or to mucosal surfaces. The immunogenic composition may comprise a plurality of active components to provide protection against disease caused by a plurality of species of transferrin receptor producing bacteria. The immunogenic compositions may further comprise an adjuvant.

In accordance with another aspect of the invention, there is provided a method for inducing protection against infection or disease caused by *Haemophilus* or other bacteria that produce transferrin receptor protein, comprising the step of administering to a susceptible host, such as a human, an effective amount of the immunogenic composition as recited above.

In accordance with another aspect of the invention, an antiserum or antibody specific for the recombinant protein, the isolated and purified Tbp1 protein or Tbp2

lysate; (d) fractionating the cell lysate to provide a first supernatant and a first pellet, the first supernatant comprising substantially a large proportion of soluble host proteins; (e) separating the first supernatant from the first pellet; (f) selectively extracting the first pellet to remove substantially all soluble host proteins and host membrane proteins therefrom to provide a second supernatant and an extracted pellet containing the inclusion bodies; (g) separating the second supernatant from the extracted pellet; (h) solubilizing the extracted pellet to provide a solubilized extract; and (i) fractionating the solubilized extract to provide a Tbp1 or Tbp2 protein containing fraction.

The cell lysate may be fractionated to provide the first supernatant and first pellet may be effected by at least one detergent extraction.

The solubilized extract may be fractionated by gel filtration to provide the Tbp1 or Tbp2 protein containing fraction, which may be subsequently dialyzed to remove at least the detergent and provide a further purified solution of Tbp1 or Tbp2 protein.

BRIEF DESCRIPTION OF DRAWINGS

The present invention will be further understood from the following description with reference to the drawings, in which:

Figure 1A shows the restriction map of two plasmid clones (pBHT1 and pBHT2) of the transferrin receptor operon of *Haemophilus influenzae* type b strain DL63.

Figure 1B shows the restriction map of clones S-4368-3-3 and JB-901-5-3 containing TfR genes from *H. influenzae* type b strain Eagan.

Figure 1C shows the restriction map of clone DS-712-1-3 containing the transferrin receptor gene from *H. influenzae* type b strain MinnA.

strain PAK 12085. Putative -35, -10 and ribosomal binding site sequences are overlined.

Figure 7 shows the nucleotide sequences of the transferrin receptor genes (SEQ ID NO: 105) and their deduced amino acid sequences (SEQ ID NO. 106 -Tbp1 and SEQ ID NO. 107 - Tbp2) from the non-typable *H. influenzae* strain SB33.

Figure 8 shows the nucleotide sequence of the Tbp2 gene (SEQ ID NO: 108) and the deduced amino acid sequence (SEQ ID NO: 109 - Tbp2) from non-typable strain *H. influenzae* strain SB12.

Figure 9 shows the nucleotide sequence of the Tbp2 gene (SEQ ID NO: 110) and the deduced amino acid sequence (SEQ ID NO: 111 - Tbp2) from non-typable strain *H. influenzae* strain SB29.

Figure 10 shows the nucleotide sequence of the Tbp2 gene (SEQ ID NO: 112) and the deduced amino acid sequence (SEQ ID NO: 113 - Tbp2) from non-typable strain *H. influenzae* strain SB30.

Figure 11 shows the nucleotide sequence of the Tbp2 gene (SEQ ID NO: 114) and the deduced amino acid sequence (SEQ ID NO: 115 - Tbp2) from non-typable strain *H. influenzae* strain SB32.

Figure 12A shows the nucleotide sequences of the promoter regions and 5'-end of the *tbp2* genes from *H. influenzae* strains Eagan (SEQ ID NO: 116), MinnA (SEQ ID NO: 117), PAK 12085 (SEQ ID NO: 118) and SB33 (SEQ ID NO: 119). The coding strand primer used to amplify *tbp2* genes by PCR is underlined (SEQ ID NO: 120).

Figure 12B shows the nucleotide sequence of the intergenic region and 5'-end of the *tbp1* genes from *H. influenzae* strains Eagan (SEQ ID NO: 121), MinnA (SEQ ID NO: 122), DL63 (SEQ ID NO: 123), PAK 12085 (SEQ ID NO: 124), SB12 (SEQ ID NO: 125), SB29 (SEQ ID NO: 126), SB30 (SEQ ID NO: 127), and SB32 (SEQ ID NO: 128). The non-

Figure 21 shows the construction scheme of plasmid JB-1600-1 which expresses *H. influenzae* strain SB12 Tbp2 from *E. coli*.

Figure 22 shows SDS-PAGE gels of products from the expression of *Haemophilus* type b Eagan Tbp1 protein, Eagan Tbp2 protein, and non-typable *H. influenzaea* SB12 Tbp2 protein from *E. coli*. Lane 1, JB-1476-2-1 (T7/Eagan Tbp1) at t_o; lane 2, JB-1476-2-1 at t=4h induction; lane 3, molecular weight markers of 200 kDa, 116 kDa, 97.4 kDa, 66 kDa, 45 kDa and 31 kDa; lane 4, JB-1437-4-1 (T7/Eagan Tbp2) at t_o; lane 5, JB-1437-4-1 at t=4h induction; lane 6, JB-1607-1-1 (T7/SB12 Tbp2) at t_o; lane 7, JB-1607-1-1 at t=4h induction.

Figure 23 shows a purification scheme for recombinant Tbp1 and Tbp2 expressed from *E. coli*.

Figure 24 shows an analysis of the purity of recombinant Tbp1 and Tbp2 purified by the scheme of Figure 23. Lane 1 contains molecular weight size markers (106, 80, 49.5, 32.5, 27.5 and 18.5 kDa), Lane 2 is *E. Coli* whole cell lysate. Lane 3 is solubilized inclusion bodies. Lane 4 is purified Tbp1 or Tbp2.

Figure 25 shows the immunogenicity of rTbp1 (upper panel) and rTbp2 (lower panel) in mice.

Figure 26 shows the reactivity of anti-Eagan rTbp1 antisera with various *H. influenzae* strains on a Western blot. Lane 1, BL21/DE3; lane 2, SB12-EDDA; lane 3, SB12 +EDDA; lane 4, SB29 - EDDA; lane 5, SB29 +EDDA; lane 6, SB33 -EDDA; lane 7, SB33 + EDDA; lane 8, Eagan -EDDA; lane 9, Eagan +EDDA; lane 10, *B. catarrhalis* 4223 - EDDA; lane 11, *B. catarrhalis* 4223 +EDDA; lane 12, *N. meningitidis* 608 - EDDA; lane 13, *N. meningitidis* 608 + EDDA; lane 14, induced JB-1476-2-1 expressing recombinant Eagan Tbp1; lane 15, molecular weight markers. Specific ~ 95 kDa bands reacted with the anti-Tbp1 antisera in lanes 3, 4, 5, 7, 8 and 9, corresponding to *H. influenzae* strains SB12, SB29, SB33 and Eagan; ~ 110 kDa bands in

a pool of the sera collected on day 27 from rabbits immunised with PV1TBP2A (rabbits 40, 41 and 42). Panel C shows the results for a pool of prebleed sera from the same, which displayed minimal specific reactivity.

In some of the above Figures, the following abbreviations have been used to designate particular site specific restriction endonucleases: R, *Eco* RI; Ps, *Pst* I; H, *Hind* III; Bg, *Bgl* II; Nde, *Nde* I; Ear, *Ear* I; and Sau, *Sau*3A I.

In Figure 28, the following abbreviations have been used to designate particular site specific restriction endonucleases: A, *Acc* I; B *Bam* HI; E, *Eco* RI; O, *Xho* I; H, *Hind* III; Ps, *Pst* I; V, *Eco* RV; X, *Xba* I, G, *Bgl* II; S, *Sal* I; K, *Kpn* I; and S*, *Sac* I.

GENERAL DESCRIPTION OF THE INVENTION

Any *Haemophilus* strain may be conveniently used to provide the purified and isolated nucleic acid which may be in the form of DNA molecules, comprising at least a portion of the nucleic acid coding for a transferrin receptor as typified by embodiments of the present invention. Such strains are generally available from clinical sources and from bacterial culture collections, such as the American Type Culture Collection.

According to an aspect of the invention, the transferrin receptor protein may be isolated from *Haemophilus* strains by the methods described by Schryvers (1989), Ogunnaviwo and Schryvers (1992) and US patent 5,141,743, the subject matter of which is hereby incorporated by reference. Although the details of an appropriate process are provided in patent US 5,141,743, a brief summary of such process is as follows. Isolation of transferrin receptor is achieved by isolating a membrane fraction from a bacterial strain expressing transferrin binding activity and purifying the transferrin receptor by an affinity method involving the sequential steps of prebinding of transferrin to the

influenzae type b strain DL63 was mechanically sheared, *Eco*RI linkers added, and a λZAP expression library constructed. The library was screened with the anti-TfR rabbit antisera and two positive clones (pBHIT1 and pBHIT2) were obtained which had overlapping restriction maps (Figure 1A and Figure 2). The clones were sequenced and two large open reading frames were identified (Figure 2). The nucleotide sequences of the transferrin receptor genes Tbp1 and Tbp2 (SEQ ID NO: 1) from *H. influenzae* DL63 and their deduced amino acid sequences (SEQ ID NO: 5 - Tbp1 and SEQ ID NO: 6 - Tbp2) are shown in Figure 3. The sequence analysis showed the TfR operon to consist of two genes (Tbp1 and Tbp2) arranged in tandem and transcribed from a single promoter (as particularly shown in Figure 2 and Figure 3). The Tbp2 protein tends to vary in molecular weight depending on the species whereas the Tbp1 protein tends to have a more consistent molecular weight with some variability across the various bacteria which have TfR genes. The molecular weight of Tbp1 is usually in the range of 94 to 106,000 whereas the molecular weight of Tbp2 varies considerably from 58 to 98 000.

Amino acid sequencing of the N-termini and cyanogen bromide fragments of transferrin receptor from *H. influenzae* DL63 was performed. The N-terminus of Tbp2 was blocked but amino acid sequences were identified by sequencing of Tbp1 and are indicated by underlining within the protein sequence of Figure 3. These peptide sequences are **Glu Thr Gln Ser Ile Lys Asp Thr Lys Glu Ala Ile Ser Ser Glu Val Asp Thr** (as shown in Figure 3, SEQ ID NO: 101) and **Leu Gln Leu Asn Leu Glu Lys Lys Ile Gln Gln Asn Trp Leu Thr His Gln Ile Ala Phe** (as shown in Figure 3; SEQ ID NO: 102). The signal sequence of Tbp1 and the putative signal sequence of Tbp2 are indicated by double overligning in Figure 3. The putative signal sequence for Tbp1 is **Met Thr Lys Lys Pro Tyr Phe Arg Leu Ser Ile**

sequences of Tbp1 and Tbp2 (SEQ ID NO: 3) and their deduced amino acid sequences (SEQ ID NO: 9 - Tbp1 and SEQ ID NO: 10 - Tbp2) from *H. influenzae* type b strain MinnA are shown in Figure 5 where the Tbp2 sequence is first in the operon. In Figure 5, Putative -35, -10 and ribosomal binding site sequences are overlined.

Chromosomal DNA from the non-typable *H. influenzae* strain PAK 12085 was prepared. The DNA was partially digested with *Sau*3A I, size-fractionated for 10-20 kb fragments, and cloned into the BamH I site of EMBL3. The library was probed with the fragments of the pBHIT clone (Figure 2) and a full-length clone encoding TfR (JB-1042-7-6) was obtained. The restriction map of clone JB-1042-7-6 is shown in Figures 1D and 2 and the nucleotide sequences of the Tbp1 and Tbp2 genes (SEQ ID NO: 4) from *H. influenzae* PAK 12085 and their deduced amino acid sequences are shown in Figure 6 (SEQ ID NOS: 11, 12), with the Tbp2 sequence first. In Figure 6, Putative -35, -10 and ribosomal binding site sequences are overlined.

Chromosomal DNA from the otitis-media derived non-typable *H. influenzae* strain SB33 was prepared. The DNA was partially digested with *Sau*3A I, size-fractionated for 10-20 kb fragments, and cloned into the BamH I site of EMBL3. The library was probed with the fragments of the pBHIT clone (Figure 2) and a full-length clone encoding TfR (JB-1031-2-9) was obtained. The restriction map of clone JB-1031-2-9 is shown in Figure 2 and the nucleotide sequences of the Tbp1 and Tbp2 genes (SEQ ID NO: 4) from *H. influenzae* SB33 and their deduced amino acid sequences are shown in Figure 7 (SEQ ID NOS: 11, 12), with the Tbp2 sequence first. The SB33 *tbp2* gene was found to have a single base deletion which resulted in a frame-shift at residue 126 and premature truncation of the resulting protein at residue 168.

SB12, SB29, SB30 and SB32 are compared in Figure 15. The Tbp2 proteins of Eagan and MinnA are identical and contain 660 amino acids, that of DL63 has 644 residues, and that of PAK 12085 has 654 residues. There is a single base deletion in the SB33 *tbp2* gene which results in a frame-shift at residue 126 and premature trunction of the resulting protein at residue 168. The missing base was confirmed by direct sequencing of PCR amplified chromosomal DNA. With the exception of Eagan and MinnA which are identical, the Tbp2 protein sequences are less conserved with only 66-70% identity, but there are several short segments of conserved sequence which can be identified in Figure 15. The PCR amplified *tbp2* genes from strains SB12, SB29, SB30 and SB32 were all found to encode full-length Tbp2 proteins. There was sequence and size heterogeneity amongst the deduced Tbp2 proteins wherein SB12 had 648 amino acids, SB29 had 631 residues, SB30 had 630 residues and SB32 had 631 residues.

Putative secondary structures of Eagan Tbp1 and Tbp2 were determined (Figures 16A and 16B). Both proteins have several transmembrane domains, with Tbp1 traversing the membrane 20 times and Tbp2 crossing it 12 times. Three exposed conserved epitopes were identified in the Tbp1 amino-terminal region (DNEVTGLGK - SEQ ID NO: 43, EQVLN/DIRDLTRYD - SEQ ID NOS: 139 and 140, and GAINEIEYENVKAVEISK - SEQ ID NO: 141) and one in the C-terminal region (GI/VYNLF/LNYRYVTWE - SEQ ID NOS: 142 and 143). Only three small conserved regions can be identified within the Tbp2 proteins of the human pathogens: CS/LGGG(G)SFD - SEQ ID NOS: 75, 144 and 145 at the N-terminal, LE/SGGFY/FGP - SEQ ID NOS: 74 and 146 located internally, and VVFGAR/K - SEQ ID NOS: 83 and 84 at the C-terminus

The discovery that the Tbp2 amino acid sequence varies between strains of *Haemophilus* allows for the grouping of *Haemophilus* into sub-groups defined by the

allows the selection of a minimal number of antigens having particular amino acid sequences (including in the form of synthetic peptides) to immunize against the disease caused by pathogens that have transferrin receptors. Such bacteria in addition to those recited above include other species of *Neisseria*, such as *Neisseria gonorrhoeae*, and *Branhamella*, including *Branhamella catarrhalis*. Such conserved amino acid sequences among many bacterial pathogens permits the generation of TfR specific antibodies, including monoclonal antibodies, that recognize most if not all transferrin receptors. Antiserum was raised against peptides corresponding to conserved portions of the transferrin receptor. This antiserum recognized the transferrin receptor in *Branhamella catarrhalis*. Such antisera are useful for the detection and neutralization of most if not all bacteria that produce TfR protein and are also useful for passive immunization against the diseases caused by such pathogens. Diagnostic assays and kits using such conserved amino acid sequences are useful to detect many if not all bacteria that produce transferrin receptor.

Epitopes containing the afore-mentioned amino acid sequences can be delivered to cells of the immune system by the use of synthetic peptides containing such sequences, or by the use of live vectors expressing such sequences, or by the direct administration of nucleic acid molecules encoding the amino acid sequence.

Some peptides containing conserved amino acid sequences within the Tbp1 proteins of *H. influenzae* type b strains Eagan, MinnA, DL63 and the nontypable strain PAK 12085 are shown in Table 2. Antibodies to some of these peptides were raised in guinea pigs (Table 4). Peptides containing conserved amino acid sequences within the Tbp2 proteins of *H. influenzae* type b strains Eagan, Minn A, DL63 and the nontypable strain PAK 12085 are

other strains, making these potentially useful diagnostic reagents (Figures 26 and 27).

Plasmids pUHIT1KFH and pUHITKFP shown in Figure 28, contain a selectable antibiotic resistance marker cloned within the transferrin receptor operon and were constructed to insertionally inactivate the transferrin receptor operon. These plasmids were used to transform *Haemophilus* to generate strains that do not produce transferrin receptor Tbp1 and/or Tbp2 as described in Example 19. Such strains are useful as negative controls (since they do not produce TfR) in *in vitro* and *in vivo* detection and diagnostic embodiments. Such strains are also expected to be attenuated for *in vivo* growth and are useful as live vaccines to provide protection against diseases caused by *Haemophilus*.

As discussed above, epitopes of transferrin receptor proteins can be delivered to cells of the immune system by the use of live vectors expressing such amino acid sequences and the live vector may be poliovirus. Referring to Figure 29 there is illustrated the construction of hybrid polioviruses expressing an epitope of transferrin receptor protein including the conserved epitope from Tbp2 LEGGFYGP (SEQ ID NO: 74). Such viruses were recognized by antibodies raised against a peptide incorporating the amino acid sequence LEGGFYGP (SEQ ID NO: 74) (Table 5) indicating that the viruses expressed this sequence in an antigenically recognisable form. PV1TBP2A and PV1TBP2B were also neutralized by rabbit antisera raised against *H. influenzae* strain DL63 *tbp2*, indicating that at least these two viruses expressed the sequence in a form recognisable to antibodies raised against the protein. All viruses were neutralisable by anti-PV1 sera, indicating that the changes in polio neutralization antigenic site I had not significantly affected other antigenic sites on the viruses. Furthermore, rabbit antiserum produced by immunization

production of *Haemophilus*-specific antisera, for vaccination against the diseases caused by species of *Haemophilus* and (for example) detecting infection by *Haemophilus*.

- peptides corresponding to portions of the transferrin receptor as typified by the embodiments described herein are advantageous as diagnostic reagents, antigens for the production of *Haemophilus*-specific antisera, for vaccination against the diseases caused by species of *Haemophilus* and (for example) for detecting infection by *Haemophilus*.

The transferrin receptor encoded by the nucleic acid molecules of the present invention, fragments and analogs thereof, and peptides containing sequences corresponding to portions of the transferrin receptor that are conserved between various isolates of *Haemophilus* and other bacteria that produce transferrin receptor, are useful in diagnosis of and immunization against diseases caused by any bacterial strain that produces transferrin receptor. In particular, peptides containing the sequences LEGGFYGP are conserved in the transferrin receptor proteins of many bacterial pathogens that produce transferrin receptor and are appropriate for diagnosis of and immunization against diseases caused by bacteria that produce transferrin receptor. Such bacteria include but are not limited to species of *Haemophilus, Neisseria* (including *N. meningitidis* and *N. gonorrhoeae*) and *Branhamella* (including *B. catarrhalis*).

It is clearly apparent to one skilled in the art, that the various embodiments of the present invention have many applications in the fields of vaccination, diagnosis, treatment of, for example, *Haemophilus* infections, and infections with other bacterial pathogens that produce transferrin receptor and the generation of immunological reagents. A further non-limiting discussion of such uses is further presented below.

immune response at mucosal surfaces. Thus, the immunogenic composition may be administered to mucosal surfaces by, for example, the nasal or oral (intragastric) routes. The immunogenic composition may be provided in combination with a targeting molecule for delivery to specific cells of the immune system or to mucosal surfaces. Some such targeting molecules include strain B12 and fragments of bacterial toxins, as described in WO 92/17167 (Biotech Australia Pty. Ltd.), and monoclonal antibodies, as described in U.S. Patent No. 5,194,254 (Barber et al). Alternatively, other modes of administration including suppositories and oral formulations may be desirable. For suppositories, binders and carriers may include, for example, polyalkalene glycols or triglycerides. Oral formulations may include normally employed incipients such as, for example, pharmaceutical grades of saccharine, cellulose and magnesium carbonate. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10-95% of the transferrin receptor, fragment analogs and/or peptides.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective, protective and immunogenic. The quantity to be administered depends on the subject to be treated, including, for example, the capacity of the individual's immune system to synthesize antibodies, and if needed, to produce a cell-mediated immune response. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner. However, suitable dosage ranges are readily determinable by one skilled in the art and may be of the order of micrograms of the transferrin receptor, analogs and fragments thereof and/or peptides. Suitable regimes for initial administration and booster doses are also

Examples of side chain modifications contemplated by the present invention include modification of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidation with methylacetimidate; acetylation with acetic anhydride; carbamylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6, trinitrobenzene sulfonic acid (TNBS); alkylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5'-phosphate followed by reduction with $NaBH_4$.

The guanidino group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2, 3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation via o-acylisourea formation followed by subsequent derivatisation, for example, to a corresponding amide.

Sulfhydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of mixed disulphides with other thiol compounds; reaction with maleimide; maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulfonic acid, phenylmercury chloride, 2-chloromercuric-4-nitrophenol and other mercurials; carbamylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphonyl halides. Tryosine residues may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

hydroxide and aluminim phosphate (collectively commonly referred to as alum) are routinely used as adjuvants in human and veterinary vaccines. The efficacy of alum in increasing antibody responses to diptheria and tetanus toxoids is will established and, more recently, a HBsAg vaccine has been adjuvanted with alum. While the usefulness of alum is well established for some applications, it has limitations. For example, alum is ineffective for influenza vaccination and inconsistently elicits a cell mediated immune response. The antibodies elicited by alum-adjuvanted antigens are mainly of the IgG1 isotype in the mouse, which may not be optimal for protection by some vaccinal agents.

A wide range of extrinsic adjuvants can provoke potent immune responses to antigens. These include saponins complexed to membrane protein antigens (immune stimulating complexes), pluronic polymers with mineral oil, killed mycobacteria and mineral oil, Freund's complete adjuvant, bacterial products, such as muramyl dipeptide (MDP) and lipopolysaccharide (LPS), as well as lipid A, and liposomes.

To efficiently induce humoral immune responses (HIR) and cell-mediated immunity (CMI), immunogens are emulsified in adjuvants. Many adjuvants are toxic, inducing granulomas, acute and chronic inflammations (Freund's complete adjuvant, FCA), cytolysis (saponins and pluronic polymers) and pyrogenicity, arthritis and anterior uveitis (LPS and MDP). Although FCA is an excellent adjuvant and widely used in research, it is not licensed for use in human or veterinary vaccines because of its toxicity.

Desirable characteristics of ideal adjuvants include:

(1) lack of toxicity;

(2) ability to stimulate a long-lasting immune response;

hepatitis B surface antigen, enhanced the host immune responses against hepatitis B virus.

Lipidation of synthetic peptides has also been used to increase their immunogenicity. Thus, Wiesmuller 1989, describes a peptide with a sequence homologous to a foot-and-mouth disease viral protein coupled to an adjuvant tripalmityl-s-glyceryl-cysteinylserylserine, being a synthetic analogue of the N-terminal part of the lipoprotein from Gram negative bacteria. Furthermore, Deres et al. 1989, reported *in vivo* priming of virus-lpecific cytotoxic T lymphocytes with synthetic lipopeptide vaccine which comprised of modified synthetic peptides derived from influenza virus nucleoprotein by linkage to a lipopeptide, N-palmityl-s-[2,3-bis(palmitylxy)-(2RS)-propyl-[R]-cysteine (TPC).

2. Immunoassays

The transferrin receptor, analogs and fragments thereof and/or peptides of the present invention are useful as immunogens, as antigens in immunoassays including enzyme-linked immunosorbent assays (ELISA), RIAs and other non-enzyme linked antibody binding assays or procedures known in the art for the detection of anti-bacterial, *Haemophilus*, TfR and/or peptide antibodies. In ELISA assays, the transferrin receptor, analogs, fragments and/or peptides corresponding to portions of TfR protein are immobilized onto a selected surface, for example a surface capable of binding proteins or peptides such as the wells of a polystyrene microtiter plate. After washing to remove incompletely adsorbed transferrin receptor, analogs, fragments and/or peptides, a nonspecific protein such as a solution of bovine serum albumin (BSA) or casein that is known to be antigenically neutral with regard to the test sample may be bound to the selected surface. This allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific

origin, the second antibody is an antibody having specificity for human immunoglobulins and in general IgG. To provide detecting means, the second antibody may have an associated activity such as an enzymatic activity that will generate, for example, a color development upon incubating with an appropriate chromogenic substrate. Quantification may then achieved by measuring the degree of color generation using, for example, a visible spectra spectrophotometer.

3. Use of Sequences as Hybridization Probes

The nucleotide sequences of the present invention, comprising the sequence of the transferrin receptor gene, now allow for the identification and cloning of the transferrin receptor genes from any species of *Haemophilus* and other bacteria that have transferrin receptor genes.

The nucleotide sequences comprising the sequence of the transferrin receptor genes of the present invention are useful for their ability to selectively form duplex molecules with complementary stretches of other TfR genes. Depending on the application, a variety of hybridization conditions may be employed to achieve varying degrees of selectivity of the probe toward the other TfR genes. For a high degree of selectivity, relatively stringent conditions are used to form the duplexes, such as low salt and/or high temperature conditions, such as provided by 0.02 M to 0.15 M NaCl at temperatures of between about 50°C to 70°C. For some applications, less stringent hybridization conditions are required such as 0.15 M to 0.9 M salt, at temperatures ranging from between about 20°C to 55°C. Hybridization conditions can also be rendered more stringent by the addition of increasing amounts of formamide, to destabilize the hybrid duplex. Thus, particular hybridization conditions can be readily manipulated, and will generally be a method of choice depending on the

type of target nucleic acid, source of nucleic acid, size of hybridization probe etc. Following washing of the hybridization surface so as to remove non-specifically bound probe molecules, specific hybridization is detected, or even quantified, by means of the label. As with the selection of peptides, it is preferred to select nucleic acid sequence portions which are conserved among species of *Haemophilus*, such as nucleic acid sequences encoding the conserved peptide sequence of Figures 8, 9, 13 and 14 and particularly listed in Tables 2 and 3. The selected probe may be at least 18 bp and may be in the range of 30 bp to 90 bp long.

4. Expression of the Transferrin Receptor Genes

Plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell may be used for the expression of the transferrin receptor genes in expression systems. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* may be transformed using pBR322 which contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR322 plasmid, or other microbial plasmid or phage must also contain, or be modified to contain, promoters which can be used by the host cell for expression of its own proteins.

In addition, phage vectors containing replicon and control sequences that are compatible with the host can be used as a transforming vector in connection with these hosts. For example, the phage in lambda GEM™-11 may be utilized in making recombinant phage vectors which can be used to transform host cells, such as *E. coli* LE392.

Promoters commonly used in recombinant DNA construction include the β-lactamase (penicillinase) and lactose promoter systems (Chang *et al.*, 1978: Itakura *et*

of this application. Samples of the deposited plasmids will become available to the public upon grant of a patent based upon this United States patent application. The invention described and claimed herein is not to be limited in scope by plasmids deposited, since the deposited embodiment is intended only as an illustration of the invention. Any equivalent or similar plasmids that encode similar or equivalent antigens as described in this application are within the scope of the invention.

Deposit Summary

Clone	ATCC Designation	Date Deposited
DS-712-1-3	75603	November 4, 1993
JB-1042-7-6	75607	November 4, 1993
JB-1424-2-8	75937	October 27, 1994
JB-1600-1	75935	October 27, 1994
JB-1468-29	75936	October 27, 1994
pT7TBP2A	75931	October 27, 1994
pT7TBP2B	75932	October 27, 1994
pT7TBP2C	75933	October 27, 1994
pT7TBP2D	75934	October 27, 1994

Strains of *Haemophilus*

Hib strain Eagan is available from Connaught Laboratories Limited, 1755 Steeles Ave. W., Willowdale, Ontario, Canada M2R 3T4.

Hib strain MinnA was obtained from the collection of Dr. Robert Munson, Department of Microbiology and Immunology, Washington University School of Medicine, Children's Hospital, St. Louis, Missouri 63110.

Hib strain DL63 was obtained from the collection of Dr. Eric Hansen, Department of Microbiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9048.

HCl, pH 8.0, centrifuged as before, resuspended in 12.5 ml of 50mM Tris-HCl, 50mM EDTA, pH 8.0, and frozen at -20°C. Then 1.25 ml of a 10 mg/ml lysozyme solution in 0.25M Tris-HCl, pH 8.0, was added to the frozen cell pellet. The pellet was thawed and incubated on ice for 45 minutes. Next, 2.5 ml of a solution of 1mg/ml proteinase K in 0.5% SDS, 0.4M EDTA, 50mM Tris-HCl, pH 7.5 was added and the mixture incubated at 50°C for 1 hour with occasional mixing. The lysate was extracted once with 15 ml of Tris-buffered phenol, then 1.5 ml of 3M sodium acetate and 30 ml of ethanol were added to precipitate the DNA. The DNA was spooled on a glass rod, then dissolved in 12.5 ml of 50mM Tris-HCl, 1mM EDTA, pH 7.5 containing 0.2 mg/ml RNAse A by rocking overnight. The sample was extracted once with an equal volume of chloroform, precipitated, and spooled as above. The DNA was dissolved in 2 ml of 50mM Tris-HCl, 1mM EDTA, pH 7.5 and stored at 4°C.

B. **Chromosomal DNA extraction from *Haemophilus influenzae* type b Eagan**

Thirty ml of culture were pelleted by centrifugation, the pellet resuspended in 25ml of TE (10mM Tris, 1mM EDTA, pH 7.5), and 2 x 5ml aliquots used for chromosomal DNA preparation. To each aliquot was added 0.6ml of 10% sarkosyl and 0.15ml of 20mg/ml proteinase K and the samples incubated at 37°C for 1 hour. The lysate was extracted once with Tris-saturated phenol and three times with chloroform:isoamyl alcohol (24:1). The aqueous phases were pooled for a final volume of 7ml. Then 0.7ml of 3M sodium acetate (pH 5.2) and 4.3 ml of isopropanol were added to precipitate the DNA which was spooled, rinsed with 70% ethanol, dried, and resuspended in 1 ml of water.

dGTP, and dTTP), and 4 μl of 5 U/μl Klenow. The mixture was incubated at 12°C for 30 minutes. 450 μl of STE (0.1M NaCl, 10mM Tris-HCl, 1mM EDTA, pH 8.0) were added, and the mixture extracted once with phenol/chloroform, and once with chloroform, before adding 1 ml of ethanol to precipitate the DNA. The sample was incubated on ice for 10 min or at -20°C overnight. The DNA was harvested by centrifugation in a microfuge for 30 minutes, washed with 70% ethanol and dried.

The DNA was resuspended in 7 μl of TE and to the solution was added 14 μl of phosphorylated *Eco* RI linkers (200 ng/μl), 3 μl of 10x ligation buffer, 3 μl of 10mM ATP, and 3μl of T4 DNA ligase (4 U/μl). The sample was incubated at 4°C overnight, then incubated at 68°C for 10 minutes to inactivate the ligase. To the mixture was added 218 μl of H_2O, 45 μl of 10x Universal buffer, and 7 μl of *Eco* RI at 30 U/μl. After incubation at 37°C for 1.5 hours, 1.5 μl of 0.5M EDTA was added, and the mixture placed on ice.

The DNA was size fractionated on a sucrose gradient, pooling fractions containing DNA of 6-10 kb. The pooled DNA was ethanol precipitated and resuspended in 5 μl of TE buffer. 200ng of insert DNA was ligated for 2-3 days at 4°C with 1 μg of ZAP II vector in a final volume of 5μl. The ligation mixture was packaged using Gigapack II Gold (Stratagene) and plated on *E. coli* SURE cells on NZY plates. The library was titrated, amplified, and stored at 4°C under 0.3% chloroform.

B. *H. influenzae* Eagan-pUC library

Chromosomal DNA prepared from *H. influenzae* Eagan by the method in Example 1C was digested with *Sau*3A I for 2, 5, and 10 minutes and samples electrophoresed on a preparative agarose gel. Gel slices which included DNA fragments between 3-10 kb in length were excised and the DNA extracted by the standard freeze-thaw procedure. Plasmid DNA from pUC 8:2 (pUC 8 with additional *Bgl* II

on *E. coli* LE392 cells. The library was titrated, then amplified and stored at 4°C under 0.3% chloroform.

Chromosomal DNA from *H. influenzae* PAK 12085 or SB33 prepared as in Example 1C was digested with *Sau*3A I (0.5 units/10 μg DNA) at 37°C for 15 minutes and size-fractionated by agarose gel electrophoresis. Gel slices corresponding to DNA fragments of 15-23 kb were excised and DNA was electroeluted overnight in dialysis tubing containing 3 ml of TAE at 14V. The DNA was precipitated twice and resuspended in water before overnight ligation with EMBL3 *BamH* I arms (Promega). The ligation mixture was packaged using the Lambda *in vitro* packaging kit (Amersham) according to the manufacturer's instructions and plated onto *E. coli* NM539 cells. The library was titrated, then amplified, and stored at 4°C in the presence of 0.3% chloroform.

Example 3

This Example illustrates screening of the libraries

A. *H. influenzae* DL63-λZAP expression library

Tbp1 and Tbp2 proteins were affinity purified on solid phase human transferrin (hTf). Briefly, a 20 ml hTf-Sepharose column was prepared according to the manufacturer's protocol for coupling protein ligands to CNBr-activated Sepharose (Sigma). The resulting matrix was washed with 3 column volumes of 50mM Tris-HCl, 1M NaCl, 6M guanidine-HCl, pH 8.0 to remove non-covalently bound hTf. The column was then equilibrated with 50mM Tris-HCl, pH 8.0 and bound hTf was iron loaded using 1 ml of 10mg/ml $FeCl_3$ in buffer containing 100mM each of sodium citrate and sodium bicarbonate, pH 8.6, followed by 2 column volumes of 50mM Tris-HCl, 1M NaCl, pH 8.0. Total bacterial membranes (300 mg total protein) were prepared from *H. influenzae* strain DL63 grown on iron deficient media as described previously (Schryvers *et al.*, 1989). Membranes were diluted to 2 mg/ml in 50mM Tris-HCl, 1M NaCl, pH 8.0 and solubilized by the addition

second round screening using the same 5'pBHIT2 probe. Second round putatives were analysed by restriction enzyme mapping and clone S-4368-3-3 (Figure 1B, Figure 2) was selected for sequence analysis.

(ii) Screening *H. influenzae* Eagan-λZAP library

The phage library was plated using standard techniques on XLI Blue cells (Stratagene) using LB plates and a 0.7% agarose overlay layer. Plaques were lifted onto nitrocellulose using standard protocols and the filters were baked at 80°C, for 2 hours, under vacuum, to fix the DNA. The 5'pBHIT2 probe of the transferrin receptor gene (Figure 2) was labelled with digoxigenin and the filters were pre-hybridized for 4 hours at 42°C, then hybridized with the labelled probe at 42°C, overnight. The filters were washed at 68°C and after autoradiography, several plaques were selected for second round screening. *In vivo* excision of phagemid DNA from second round putatives was performed according to protocols provided with the λZAP system (Promega). Four clones with identical ~2.5 kb *Eco* RI inserts were obtained of which JB-901-5-3 in Figure B, Figure 2 is an example. Putative plaques were also amplified and phage DNA was purified from 500 ml of culture. Insert DNA was excised by digestion with *Xba* I and was cloned into pUC 8:2 (pUC 8 containing additional *Bgl* II and *Xba* I sites in its multiple cloning site) which had been digested with *Xba* I and dephosphorylated. Clone JB-911-3-2 (Figure 17) contains the 3'-half of the *H. influenzae* Eagan TfR operon.

(iii) Screening EMBL 3 libraries

The *H. influenzae* MinnA library was plated onto LE392 cells on NZCYM plates using 0.7% top agarose in NZCYM as overlay. Plaque lifts onto nitrocellulose filters were performed following standard procedures, and filters were processed and probed with the 5'pBHIT2 probe (Figure 2) labelled with digoxigenin. Putative plaques

the manufactures recommendations. Samples were sequenced using the ABI model 370A DNA Sequencer and dye terminator chemistry according to manufacturers' protocols. The sequence of the TfR operon from strain DL63 is illustrated in Figure 3, that of strain Eagan in Figure 4, that of strain MinnA in Figure 5, that of PAK 12085 in Figure 6 and that of SB33 in Figure 7.

Example 5

This Example illustrates the PCR amplification of the *tbp2* genes from non-typable *H. influenzae* strains SB12, SB29, SB30, and SB32.

Chromosomal DNA from non-typable *H. influenzae* strains SB12, SB29, SB30, and SB32 was prepared as described aobve. The TfR genes are arranged as an operon with *tbp2* followed by *tbp1* (see Figures 12A and 12B). Oligonucleotides were synthesized to the 5'-end of the *tbp2* and the reverse complement of the 5'-end of the *tbp1* coding sequences. The primers were: GGATCCATATGAAATCTGT ACCTCTTATCTCTGGT (SEQ ID NO: 120) corresponding to MKSVPLISGS (SEQ ID NO: 147) from the leader sequence of Tbp2 and TCTAGAAGCTTTTTTAGTCATTTTTAGTATTCCAT (SEQ ID NO: 137) which is the reverse complement of the leader sequence MTKK (SEQ ID NO: 138) of Tbp1 and a part of the intergenic sequence (Figures 12A and 12B). PCR amplification was performed in buffer containing 10mM Tris/HCl pH 8.3, 50 mM potassium chloride and 1.5 mM magnesium chloride. Each 100 μl reaction mixture contained 5 ng of chromosomal DNA, 1 μg of each primer, 5 units amplitaq DNA polymerase (Perkin Elmer Cetus) and 0.4 mM dNTPs (Perkin Elmer Cetus). The cycling conditions were 25 cycles of 94°C for 1.0 min, 45°C for 2.0 min and 72°C for 1.5 min. Specific 2 kb fragments were amplified for each sample (Figure 13). SB33 DNA was used as a positive control (Lane 1). Chromosomal DNA used for amplification of the Tbp2 gene were lane 1, SB33; lane 2, SB12; lane 3, SB29; lane 4, SB30; and lane

type b Eagan. The predicted secondary structures depicted in Figures 16A and 16B were arrived at using the procedures described above. However, the inventors have not yet been able to verify that the secondary structure is accurately depicted by these Figures.

Conserved epitopes of Tbp1 and Tbp2 proteins from several different bacteria were identified by sequence alignment as shown in Figures 14 and 15 respectively. Some such conserved epitopes include:

TBP1	DNEVTGLGK	SEQ ID NO:43
TBP1	EQVLNIRLTRYDPGI	SEQ ID NO:44
TBP1	GAINEIEYENVKAVEISKG	SEQ ID NO:45
TBP1	GALAGSV	SEQ ID NO:46
TBP2	LEGGFYGP	SEQ ID NO:74
TBP2	CSGGGSFD	SEQ ID NO:75
TBP2	YVYSGL	SEQ ID NO:76
TBP2	CCSNLSYVKFG	SEQ ID NO:77
TBP2	FLLGHRT	SEQ ID NO:78
TBP2	EFNVDF	SEQ ID NO:79
TBP2	NAFTGTA	SEQ ID NO:80
TBP2	VNGAFYG	SEQ ID NO:81
TBP2	LEGGYF	SEQ ID NO:82
TBP2	VVFGAR	SEQ ID NO:83

Furthermore, in combination with the predicted secondary structures, four conserved exposed epitopes were identified on Tbp1 and two were identified on Tbp2. These are:

Tbp1	DNEVTGLGK	SEQ ID NO:43
Tbp1	EQVLN/DIRDLTRYD	SEQ ID NOS: 139 and 140
Tbp1	GAINEIEYENVKAVEISK	SEQ ID NO:141
Tbp1	GI/VYNLF/LNYRYVTWE	SEQ ID NOS:142 and 143
Tbp2	CS/LGGG(G)SFD	SEQ ID NOS: 75, 144 and 145
Tbp2	LE/SGGFY/FGP	SEQ ID NOS: 74 and 146

Proteins, polypeptides or peptides containing the afore-mentioned conserved amino acid sequences are particularly useful as detecting means in diagnostic

Example 8

This Example illustrates the construction of plasmid JB-1424-2-8 which expresses Eagan Tbp2 from *E. coli*.

Referring to Figure 18, there is shown plasmid S-4368-3-3 which contains the entire *tbp2* gene from *H. influenzae* type b Eagan. Figure 18 illustrates plasmid JB-1424-2-8 and Figure 19 shows the oligonucleotides used. Plasmid JB-1424-2-8 was introduced into *E. coli* strain BL21/DE3 by electroporation to generate *E. coli* strain JB-1437-4-1. Upon induction with IPTG or lactose, Tbp2 was expressed by *E. coli* JB-1437-4-1 as shown in Figure 22. Lane 1, JB-1476-2-1 (T7/Eagan Tbp1) at t_o; lane 2, JB-1476-2-1 at t=4h induction; lane 3, molecular weight markers of 200 kDa, 116 kDa, 97.4 kDa, 66 kDa, 45 kDa and 31 kDa; lane 4, JB-1437-4-1 (T7/Eagan Tbp2) at t_o; lane 5, JB-1437-4-1 at t=4h induction; lane 6, JB-1607-1-1 (T7/SB12 Tbp2) at t_o; lane 7, JB-1607-1-1 at t=4h induction.

Example 9

This Example illustrates the construction of plasmids which encode a lipoprotein leader sequence before the Tbp2 sequence.

Oligonucleotides used for the construction of plasmids with lipoprotein leader sequences derived from *E. coli* lpp (SEQ ID NOS: 88 and 89), rlpB (SEQ ID NOS: 90 and 91), and pal (SEQ ID NOS: 92 and 93) preceeding Tbp2 are shown in Figure 20. Plasmids constructed and corresponding strains generated are illustrated in Table 1 below.

Example 10

This Example illustrates the construction of plasmid JB-1600-1 which expresses SB12 Tbp2 from *E. coli*.

Plasmid DS-1047-1-2 (Figure 21) contains the PCR-amplified SB12 *tbp2* gene. The *tbp2* gene was excised as a Nde I to EcoR I restriction fragment and inserted into the pT7-7 expression vector to generate plasmid JB-1600-

fractions were analysed by SDS PAGE and those containing purified Tbp1 or Tbp2 were dialysed overnight at 4°C against 50 mM Tris, pH 8.0 and then centrifuged at 20,000 x *g* for 30 min. The protein remained soluble under these conditions and the purified Tbp1 and Tbp2 were stored at -20°C.

The SDS-PAGE analysis of the purification process is shown in Figure 24. Lanes 1, prestained molecular weight protein markers (106, 80, 49.5, 32.5, 27.5, 18.5 kDa); lanes 2, *E.coli* whole cell lysates; lanes 3, solubilized inclusion bodies; lanes 4, purified Tbp1 or Tbp2.

Example 12

This Example illustrates immunogenicity studies of recombinant Tbp1 and Tbp2 in mice.

Groups of five Balb/c mice were injected subcutaneously (s.c.) on days 1, 29 and 43 with purified rTbp1 or rTbp2 (1 μg to 10 μg), prepared as described in Example 11, in the presence of $AlPO_4$ (1.5 mg per dose). Blood samples were taken on days 14, 28, 42 and 56 for analysis of the anti-rTbp1 and anti-rTbp2 antibody titers by EIA. The results of the immunogenicity studies are illustrated in Figure 25.

Example 13

This Example illustrates the development of EIAs for determination of anti-rTbp1 and anti-rTbp2 antibodies in mouse sera.

Anti-rTbp1 and anti-rTbp2 antibody titres were determined essentially as described by Panezutti *et al.* (1993). Microtiter wells were coated with 0.5 μg of rTbp1 or rTbp2, prepared as described in Example 11, for 16 h at room temperature, then blocked with 0.1% (w/v) BSA in PBS. The sera were serially diluted, added to the wells, then incubated for one hour at room temperature. Affinity-purified $F(ab')_2$ fragments of goat anti-mouse IgG (Fc specific) antibody conjugated to horseradish peroxidase were used as second antibody. The reactions

recombinant Eagan Tbp2, with various strains of *H. influenzae*.

Whole cell lysates of *H. influenzae* strains grown in BHI media supplemented with NAD and heme (Harkness et al., 1992) ± EDDA were separated on an SDS PAGE gel, transferred to nitrocellulose membrane, and probed with guinea pig anti-Tbp2 antisera raised to purified recombinant Eagan Tbp2 (Figure 27). Lane 1, molecular weight markers; lane 2, induced JB-1437-4-1 expressing recombinant Eagan Tbp2; lane 3, SB12-EDDA; lane 4, SB12 +EDDA; lane 5, SB29 -EDDA; lane 6, SB29 +EDDA; lane 7, SB30 -EDDA; lane 8, SB30 +EDDA; lane 9, SB32 -EDDA; lane 10, SB33-EDDA; lane 11, SB33 +EDDA; lane 12, PAK -EDDA; lane 13, PAK +EDDA; lane 14, Eagan -EDDA; lane 15, Eagan +EDDA. Specific bands of about 60-70 kDa were reactive with the anti-Tbp2 antisera in lanes 3, 6, 7, 8, 13, 14 and 15, corresponding to *Haemophilus* strains SB12, SB29, SB30, PAK and Eagan.

Example 16

This Example illustrates the synthesis of synthetic peptides corresponding to conserved regions in Tbp2 and Tbp1.

The deduced amino acid sequences of Tbp1 and Tbp2 were compared as shown in Figures 14 and 15 respectively. This comparison identified regions of amino acid sequence conservation within the transferrin receptor described above and, as shown in Tables 2 and 3, peptides were synthesized containing a portion of the transferrin receptor. Such synthesis may be effected by expression in a suitable host of recombinant vectors containing nucleic acid encoding said peptides or by standard peptide synthesis.

Briefly, peptides were synthesized using an ABI 430A peptide synthesizer and optimized t-Boc chemistry using the conditions recommended by the manufacturer, and peptides were cleaved from the resin using hydrofluoric

washing buffer. The plates were developed using the substrate tetramethylbenzidine (TMB) in H_2O_2 (ADI, Toronto), the reaction was stopped with 1N H_2SO_4 and the optical density was measured at 450 nm using a Titretek Multiskan II (Flow Labs., Virginia). Two irrelevant peptides of 32 amino acid residues were included as negative controls in these ELISAs. Assays were performed in triplicate, and the reactive titer of each antiserum was defined as the dilution consistently showing a 2-fold increase in absorbance value over those obtained from the negative controls. The antisera raised in guinea pigs were monospecific for the peptide used for immunization. The titres of the sera obtained following immunization are shown in Table 4.

Peptides of the present invention comprise single copies of any of those shown in Tables 2 and 3 or peptides containing multiple copies of analogs thereof. A peptide may further comprise multiples of different peptides selected from those shown in Tables 2 and 3 or analogs thereof and include suitable carrier molecules. It is preferred that the peptides from conserved regions be used to develop antibodies because an immuno- or other type of binding assay can then be used to detect several species of *Haemophilus*. Tables 2 and 3 therefore set out several other conserved regions of transferrin receptor to identify other peptides which would be useful in diagnosis, immunization and medical treatment.

Example 18

This Example describes the ability of antiserum raised against peptides corresponding to conserved portions of transferrin receptor to recognize the transferrin receptor of *Branhamella catarrhalis*.

Guinea pigs were immunized with peptide, corresponding to conserved portions of transferrin receptor, and antisera obtained are described in Example 17. A whole-cell extract of *Branhamella catarrhalis* was

Example 19

This Example illustrates the generation of *H. influenzae* strains that do not produce transferrin receptor.

A 2.55 *Eco* RI fragment of the insert from pBHIT1 was subcloned into the *Eco* RI site of pUC4K, resulting in removal of the Tn903 kanamycin resistance (kan) cassette from this vector (pUHIT1; Figure 28). This subcloning step facilitated the subsequent insertion of either a HincII or PstI pUC4K fragment containing the kan cassette into the *Hind* III or *Pst* I sites of pUHIT1 as both are unique sites in this construction to produce pUHIT1KFH and pUHIT1KFP, Figure 28. Following digestion with *Eco* RI to remove the interrupted gene sequences, the constructs were introduced into the *H. influenzae* wild type genome by transformation using M-IV media as described previously (Barcak *et al.*, 1991) and transformants were selected on BHINH agar containing 20 μg/ml kanamycin.

Example 20

This Example illustrates the construction of polioviruses expressing an epitope of a transferrin receptor.

A cDNA clone of bases 1175 to 2956 of the poliovirus type 1, Mahoney strain (PV1-M) genome was cut with restriction enzymes *Sau* I and *Hind* III. These enzymes excise a fragment containing bases 2754 to 2786, which encodes PV1-M amino acids 1094 to 1102, as shown in Figure 29. In this application, we use the four-digit code for poliovirus amino-acids; for example, 1095 is amino acid 95 of capsid protein VP1. New hybrid cDNA clones encoding both poliovirus and transferrin receptor amino-acid sequences were constructed by replacing the excised fragment with synthetic oligonucleotides coding for amino acids from *H. influenzae* Tbp2. The new hybrid cDNA clones were cut with restriction enzymes Nhe I and

these two viruses expressed the sequence in a form recognisable to antibodies raised against the protein. All viruses were neutralisable by anti-PV1 sera, indicating that the changes in polio neutralization antigenic site I had not significantly affected other antigenic sites on the viruses.

Example 21

This Example illustrates the use of poliovirus hybrids to induce high titer antisera against Tbp2.

Rabbits were inoculated with CsCl-purified PV1TBP2A (rabbits #40, 41, 42). Note that, although the viruses used were live, poliovirus does not replicate in rabbits and that any response observed is effectively the response to an inactivated antigen. On day 1, rabbits were inoculated with 1 ug of virus in Freund's complete adjuvant subcutaneously on the back, and, on day 14, the rabbits were boosted with 1 ug of virus in Freund's incomplete adjuvant inoculated subcutaneously on the back. The rabbits were bled on day 0 (prebleed) and on day 27. The dose of virus per inoculation was 2.5×10^8 pfu, which was determined from A_{260} values to be approximately 3.0×10^{11} virions. This equivalent to 0.5 pmol of virus or 30 pmol of the LEGGFYG (SEQ ID NO: 74) epitope, since each virion expresses 60 copies of the epitope.

SUMMARY OF THE DISCLOSURE

In summary of this disclosure, the present invention provides purified and isolated DNA molecules containing transferrin receptor genes, the sequences of these transferrin receptor genes and the derived amino acid sequences thereof. The invention also provides peptides corresponding to portions of the transferrin receptor. The genes, DNA sequences, recombinant proteins and peptides are useful for diagnosis, immunization and the generation of diagnostic and immunological reagents. Vaccines based upon expressed recombinant Tbp1 and/or

TABLE 1

leader	1st residue	plasmid	strain
E. coli lpp	Cys	JB-1360-1R-10	JB-1407-1-1
E. coli lpp	Ser	JB-1366-1R-7	JB-1407-3-1
E. coli pal	Cys	JB-1360-3-10	JB-1407-2-1
E. coli pal	Ser	JB-1366-3R-5	JB-1407-4-4
E. coli rlpB	Cys	JB-1399-1	JB-1437-1-1
E. coli rlpB	Ser	JB-1378-7	JB-1407-5-1

TABLE 2 (cont)

TBP1-28	794-829	NELLGKRALGNNSRNVKSTRKLTRAWHILDVSGYYM	40
TBP1-29	825-854	SGYYMVNRSILFRLGVYNLLNYRYVTWEAV	41
TBP1-30	843-865	LLNYRYVTWEAVRQTAQGAEFDI	42
TBP1-31	42-50	DNEVTGLGK	43
TBP1-32	61-76	EQVLNIRDLTRYDPGI	44
TBP1-33	61-95	EQVLNIRDLTRYDPGISVVEQGRGASSGYSIRGMD	45
TBP1-34	128-146	GAINEIEYENVKAVEISKG	46
TBP1-35	155-161	GALAGSV	47
TBP1-1	1-14	AETQSIKDTKEAISC[2]	48

1. Residue number from the sequence of Tbp1 of *H. influenzae* type b strain Eagan (as shown in Figure 8).

2. Cysteine added to facilitate coupling to a carrier protein, for example KLH.

Table 3 (Cont)

TBP2-27	130-134	YVYSGL	76
TBP2-28	345-355	CCSNLSYVKFG	77
TBP2-29	401-407	FLLGHRT	78
TBP2-30	450-456	EFNVDF	79
TBP2-31	485-491	NAFTGTA	80
TBP2-32	516-522	VNGAFYG	81
TBP2-33	527-532	ELGGYF	82
TBP2-34	562-566	VVFGAR	83
TBP2-35	562-568	VVFGAK	84
TBP2-36	231-238	LEGGFYG	85

1. Residue number from the sequence of Tbp2 of *H. influenzae* type B Eagan strain (as shown in Figure 9).

TABLE 5

Neutralizing activity of anti-Tbp2 and anti-peptide sera against polio/Tbp2 hybrid viruses

Sera [a]	Neutralizing Titre v. Virus [b]				
	PV1TBP2A	PV1TBP2B	PV1TBP2C	PV1TBP2D	PV1XLD
Rb @PV1	>40,9600	25,844	20,480	16,763	>40,960
Rb 516 D0	<4	<4	<4	<4	<4
Rb 516 D42	40	20	<4	<4	<4
GP561, 562 D0 pool	<4	<4	<4	<4	<4
GP 561 D56	>2048	>2048	>2048	1164	<4
GP 562 D56	>2048	>2048	25	10	<4
GP558, 559, 560 D56 pool	<4	<4	<4	<4	<4

[a] Rb @PV1 is pool of rabbit immune sera raised against PV1XLD. Rabbit 516 was immunised with three successive 3 μg dses of recombinant *H. influenzae* DL63 transferrin binding protein 2 on days 1, 14 and 28. Serum was collected on days 0 (D0) and 42 (D42). Guinea-pigs were immunized with four successive doses of 200μg of peptide on days 1, 14, 28 and 42. Sera were collected on day 0 (D0) and day 56 (D56). Guinea-Pigs 561 and 562 received a peptide containing the sequence LEGGFYGP (SEQ ID NO:74). Guinea-pigs 558, 559 and 556 received a control peptide with an unrelated sequence.

[b] Titre is the inverse dilution of serum giving a 50% endpoint in a virus neutralization assay versus 100 $TCID_{50}$ of virus.

LIST OF REFERENCES

Barcak *et al.*, (1991) Methods Enzymol. **204**: 321-342.

Berkowitz *et al.*, (1987) J. Pediatr. **110**:509.

Black *et al.*, (1991) Pediatr. Infect. Dis. J. **10**:97.

Bluestone, N. (1982) Engl. J. Med. **306**:1399.

Chang *et al.*, (1978) Nature **375**:615.

Chou, *et al.*, (1978). Annual Reviews of Biochemistry **47**, 251-276.

Claesson *et al.*, (1989) J. Pediatr. **114**:97.

Cornelissen *et al.*, (1992) J. Bacteriol. **174**:5788

Danve, *et al.*, (1993). Vaccine **11**, 1214-1220.

Deres *et al.*, (1989) Nature **342**:651.

Gerlach, *et al.*, (1992) Infect. Immun. **608**:325

Goeddel *et al.*, (1979) Nature **281**:544.

Goeddel *et al.*, (1980) Nucl. Acids Res. **8**:4057

Harkness *et al.*, (1992) J. Bacteriol. **174**:2425.

Holland *et al.*, (1992) Infect. Immun. **60**:2986.

Hopp, T.P. (1986) Journal of Immunological Methods **88**, 1-18.

Itakura *et al.*, (1977) Science **198**:1056.

Jarosik *et al.*, (1994). Infection and Immunity **62**, 2470-2477.

Legrain *et al.*, (1993). Gene **130**:73

Lockhoff *et al.*, (1991) Chem. Int. Ed. Engl. **30**:1611.

Mickelsen and Sparling, (1981) Infect. Immun. **33**:555.

Morton *et al.*, (1993) Infection and Immunity **61**, 4033-4037.

Murdin *et al.*, (1992) J. Gen. Viral **73**: 607.

Murdin *et al.*, (1991) Microbial Pathogenesis **10**:27.

Nixon-George *et al.*, (1990) J. Immunol. **14**:4798.

Ogunnariwo, and Schryvers, (1992) Avian Dis. **36**:655.

CLAIMS

What we claim is:

1. A purified and isolated nucleic acid molecule encoding a transferrin receptor protein of a strain of *Haemophilus* or a fragment or an analog of the transferrin receptor protein.

2. The nucleic acid molecule of claim 1, wherein the strain of *Haemophilus* is a strain of *Haemophilus influenzae*.

3. The nucleic acid molecule of claim 2, wherein the strain of *Haemophilus influenzae* is a strain of *Haemophilus influenzae* type b.

4. The nucleic acid molecule of claim 3, wherein the strain of *Haemophilus influenzae* type b is selected from the group consisting of DL63, MinnA and Eagan.

5. The nucleic acid molecule of claim 2, wherein the strain of *Haemophilus influenzae* is a non-typable *Haemophilus influenzae* strain.

6. The nucleic acid molecule of claim 5, wherein the strain of non-typable *Haemophilus influenzae* is selected from the group consisting of PAK 12085, SB12, SB29, SB30, SB32 and SB33.

7. The nucleic acid molecule of claim 1 encoding only the Tbp1 protein of the *Haemophilus* strain.

8. The nucleic acid molecule of claim 1 encoding only the Tbp2 protein of the *Haemophilus* strain.

9. The nucleic acid molecule of claim 1 encoding a fragment of the transferrin receptor protein of a strain of *Haemophilus* having a conserved amino acid sequence which is conserved among bacteria that produce transferrin receptor protein.

10. The nucleic acid molecule of claim 9, wherein the conserved amino acid sequence has an amino acid sequence contained within the amino acid sequences of the peptides shown in Tables 2 and 3 for *Haemophilus influenzae* type

or the fragment or the analog of the transferrin receptor.

17. The expression vector of claim 16, wherein the nucleic acid molecule encodes substantially all of the transferrin receptor protein of the *Haemophilus strain*.

18. The expression vector of claim 16, wherein the nucleic acid molecule encodes only the Tbp1 or only the Tbp2 protein of the *Haemophilus* strain.

19. The expression vector of claim 16, wherein the expression means includes a nucleic acid portion encoding a leader sequence for secretion from the host of the transferrin receptor protein or the fragment or the analog of the transferrin receptor protein.

20. The expression vector of claim 16, wherein the expression means includes a nucleic acid portion encoding a lipidation signal for expression from the host of a lipidated form of the transferrin receptor protein or the fragment or the analog of the transferrin receptor protein.

21. The expression vector of claim 16 having the identifying characteristics of plasmid JB-1424-2-8 having ATCC Accession No. 75937, JB-1600-1 having ATCC Accession No. 75935 or JB-1468-29 having ATCC Accession No. 75936.

22. A transformed host containing an expression vector as claimed in claim 16.

23. The host of claim 22 which is selected from the group consisting of JB-1476-2-1, JB-1437-4-1 and JB-1607-1-1.

24. A recombinant transferrin receptor protein or fragment or analog thereof producible by the transformed host of claim 22.

25. An isolated and purified Tbp1 protein of a strain of *Haemophilus* free from the Tbp2 protein of the *Haemophilus* strain.

36. The peptide of claim 29 comprising an amino acid sequence which is conserved among bacteria that produce transferrin receptor protein.

37. The peptide of claim 36 comprising an amino acid sequence which is conserved among strains of *Haemophilus*.

38. The peptide of claim 36, wherein the peptide includes an amino acid sequence LEGGFYGP (SEQ ID NO: 74) or LEGGFYG (SEQ ID NO: 85).

39. The peptide of claim 29 having an amino acid sequence selected from those presented in Table 2 or Table 3 for the Eagan strain of *Haemophilus influenzae* type b and the corresponding amino acid sequences of other strains of *Haemophilus influenzae*.

40. An immunogenic composition, comprising at least one active component selected from the group consisting of:

(A) a purified and isolated nucleic acid molecule encoding a transferrin receptor protein of a strain of *Haemophilus* or a fragment or an analog of the transferrin receptor protein;

(B) a purified and isolated nucleic acid molecule having a DNA sequence selected from the group consisting of:

(a) any one of the DNA sequences set out in Figure 3, 4, 5, 6, 7, 8, 9, 10 or 11 (SEQ ID NOS: 1, 2, 3, 4, 105, 108, 110, 112, 114) or the complementary DNA sequence of any one of said sequences;

(b) a DNA sequence encoding one of the amino acid sequences set out in Figure 3, 4, 5, 6, 7, 8, 9, 10 or 11 (SEQ ID NOS: 5, 6, 7, 8, 9, 10, 11, 12, 106, 107, 109, 111, 113, 115) or the complementary DNA sequence thereto; and

(c) a DNA sequence which hybridizes under stringent conditions to any one of the DNA sequences defined in (a) or (b);

(C) a recombinant transferrin receptor protein or fragment or analog thereof producible is a transformed

against disease caused by a plurality of species of transferrin receptor producing bacteria.

45. The immunogenic composition of claim 40 further comprising an adjuvant.

46. A method for inducing protection against disease caused by a bacterial pathogen that produces transferrin receptor, comprising administering to a susceptible host an effective amount of the immunogenic composition of claim 40.

47. The method of claim 46, wherein the bacterial pathogen is a *Haemophilus* bacterium.

48. The method of claim 46, wherein the susceptible host is a human.

49. The method of claim 46, wherein said immunogenic composition is that of claim 44.

50. An antiserum or antibody specific for a recombinant protein as claimed in claim 24, an isolated and purified protein of claim 25 or 26, a synthetic peptide as claimed in claim 29 or an immunogenic composition as claimed in claim 40.

51. A live vector for delivery of transferrin receptor to a host, comprising a vector containing the nucleic acid molecule of claim 1 or 12.

52. The live vector of claim 51, wherein the vector is selected from *Salmonella*, BCG, adenovirus, poxvirus, vaccinia and poliovirus.

53. The live vector of claim 51, wherein the vector is poliovirus and the nucleic acid molecule codes for a fragment of transferrin receptor having an amino acid sequence of LEGGFYGP (SEQ ID NO: 74) or LEGGFYG (SEQ ID NO: 85).

54. A plasmid vector having the identifying characteristics of pT7TBP2A having ATCC Accession No. 75931, pT7TBP2B having ATCC Accession No. 75932, pT7TBP2C having ATCC Accession No. 75933 or pT7TBP2D having ATCC Accession No. 75934.

61. The method of claim 60 wherein the cell lysate is fractionated by centrifugation thereof.

62. The method of claim 61 wherein the step of selectively extracting the first pellet comprises at least one detergent extraction.

63. The method of claim 62 wherein the solubilized extract is fractionated by gel filtration to provide said Tbp1 or Tbp2 protein containing fraction.

64. The method of claim 63 including subsequently dialyzing the Tbp1 or Tbp2 protein containing fraction to remove at least said detergent to provide a further purified solution of Tbp1 or Tbp2 protein.

65. The method of claim 60 wherein said strain of *Haemophilus* is a strain of *Haemophilus influenzae*.

66. The host of claim 22 wherein said host is a *Haemophilus* strain genetically modified by said expression vector.

INTERNATIONAL SEARCH REPORT

Intern al Application No
PCT/CA 94/00616

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/285 C12N1/21 A61K39/395 C07K16/12
//(C12N1/21,C12R1:21)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	INFECTION AND IMMUNITY, vol.60, no.7, July 1992, WASHINGTON US pages 2986 - 2991 HOLLAND, J. ET AL.; 'Evidence for in vivo expression of transferrin-binding proteins in haemphilus influenzae type b' see the whole document ---	1-4, 7-10,14, 25-28, 40-50, 60-65
X	MICROBIAL PATHOGENESIS, vol.14, May 1993 pages 389 - 398 GRAY-OWEN, S.D. ET AL.; 'The interaction of primate transferrins with receptors on bacteria pathogenic to humans' see the whole document ---	1-4, 7-10,14, 25-28, 40-50, 60-65
	-/--	

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

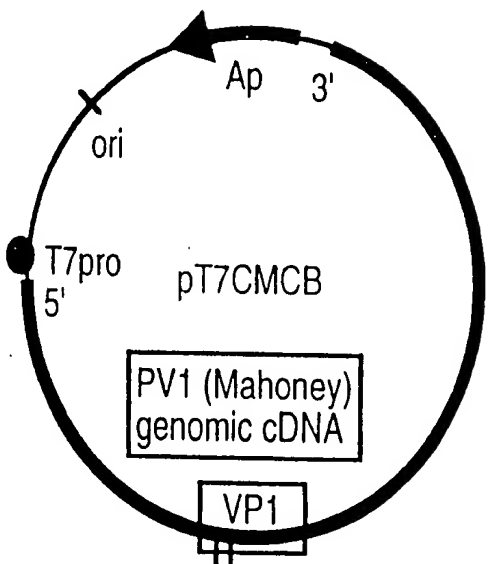
* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 February 1995	28. 02. 95

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Nauche, S

Form PCT/ISA/210 (second sheet) (July 1992)



Sal I

Hind III

nt 2753 G|TC GAC AAC CCA GCT TCC ACC ACC AAT AAG GAC A|AG CTT 2791

CAG CT|G TTG GGT CGA AGG TGG TGG TTA TTC CTG TTC GA| A

aa1094 V D [N P A S T T N K D] K L 1102

Clone	Sequence encoded	SEQ. ID NO:	Viable virus recoverd/ strain designation
pT7XLD	NPASTTNKD	132	Yes/PV1-XLD
pT7TBP2A	NPASTTSLEGGFYGPKD	133	Yes/PV1TBP2A
pT7TBP2B	NPASTTSLEGGFYGKD	134	Yes/PV1TBP2B
pT7TBP2C	NPASTTLEGGFYGPKD	135	Yes/PV1TBP2C
pT7TBP2D	NPASTTLEGGFYGKD	136	Yes/PV1TBP2B

FIG.29.

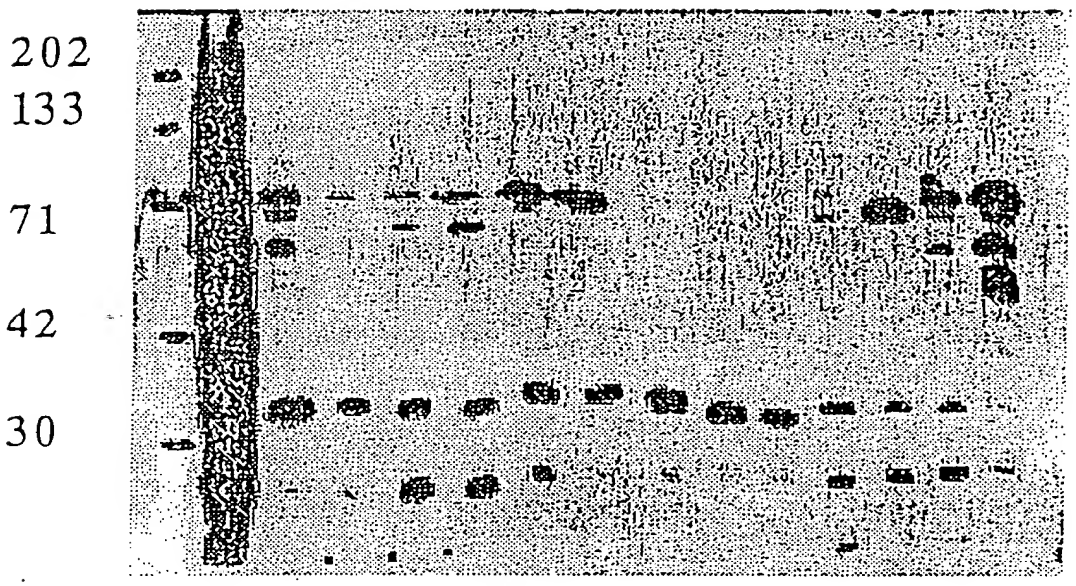


FIG. 27.

Kinetics of Antibody Response to TBP1/TBP2 in Mice

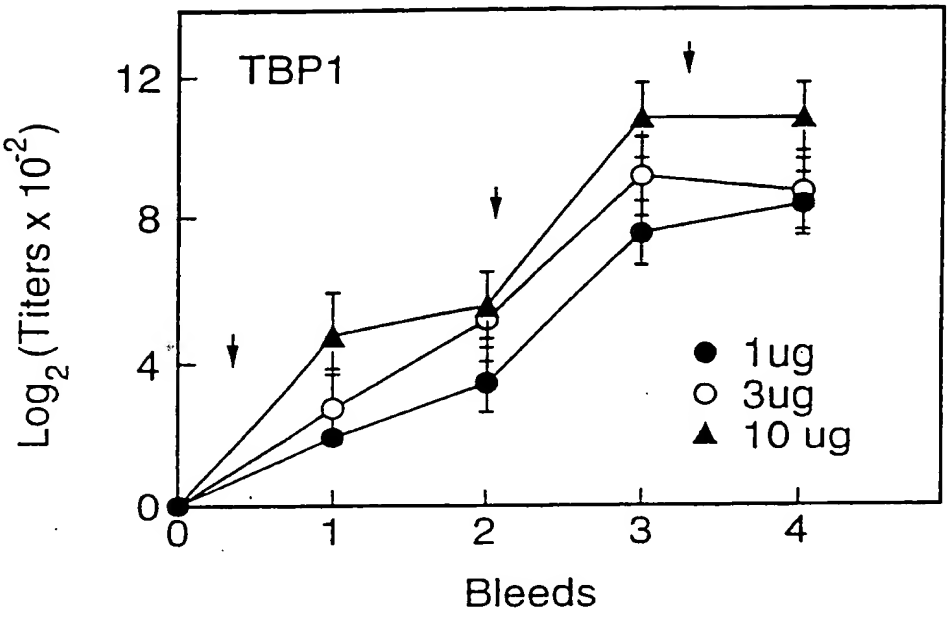


FIG.25A.

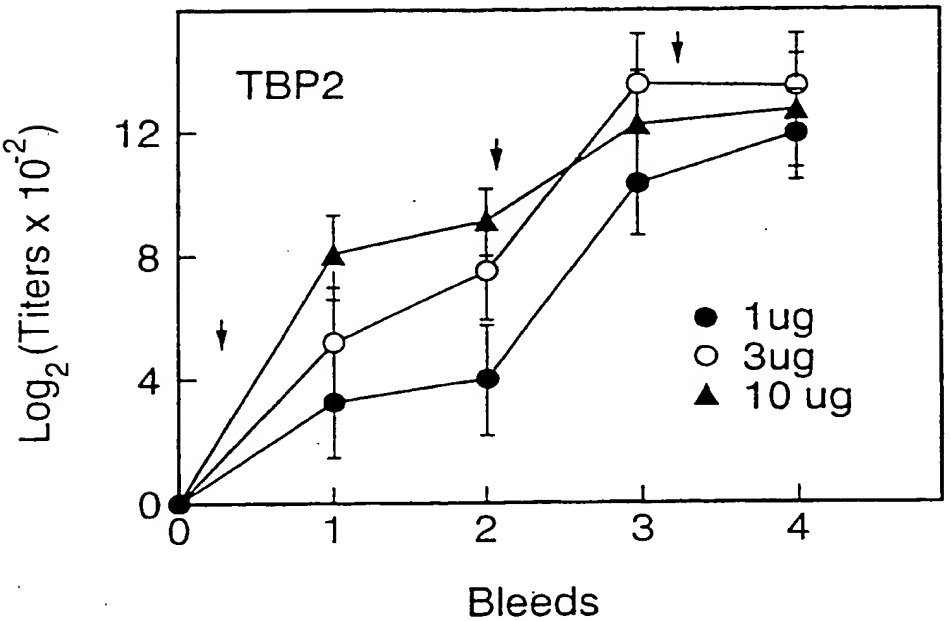


FIG.25B.

PURFICATION OF rTBP1/ rTBP2 FROM *E. COLI*

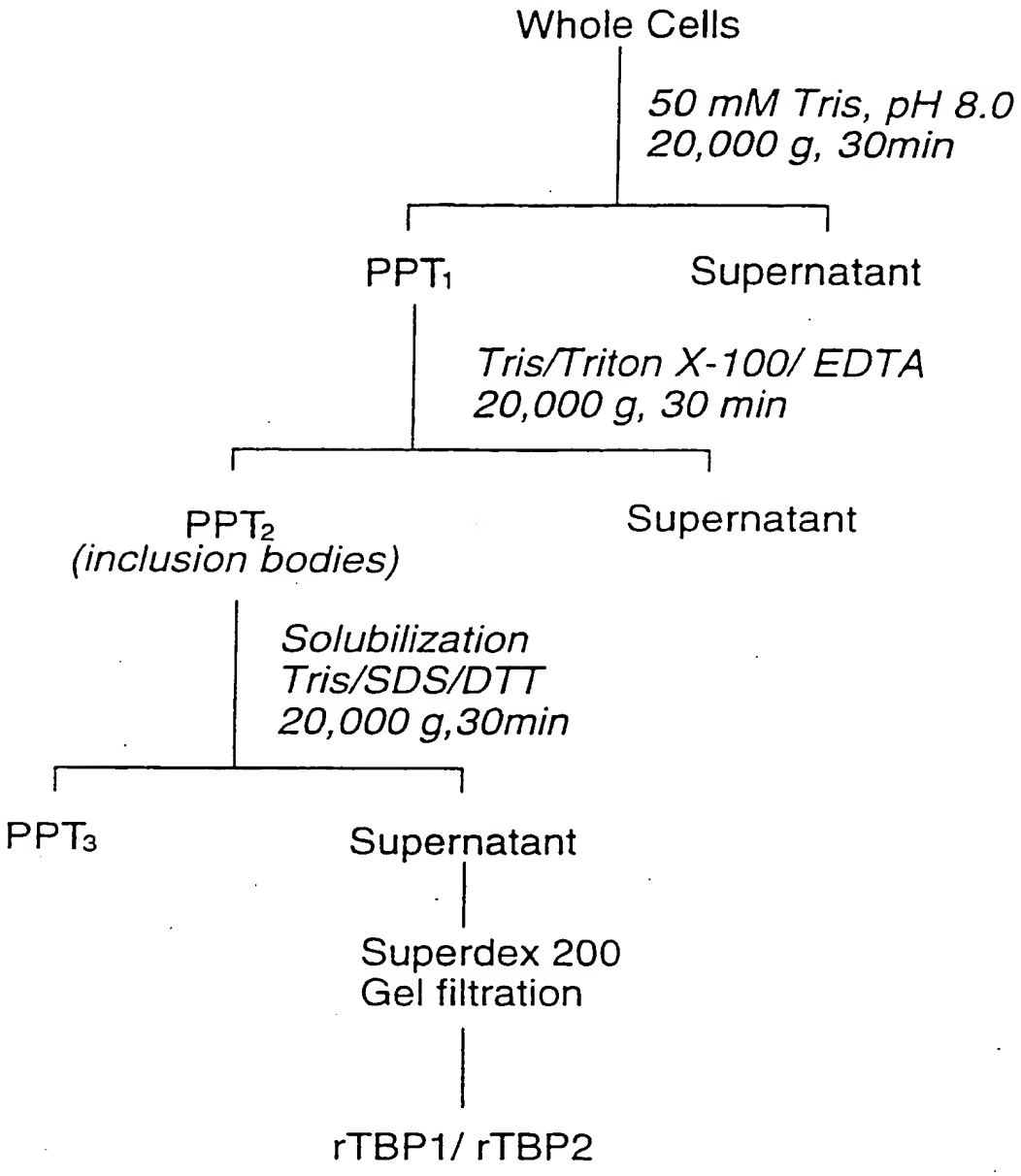


FIG.23.

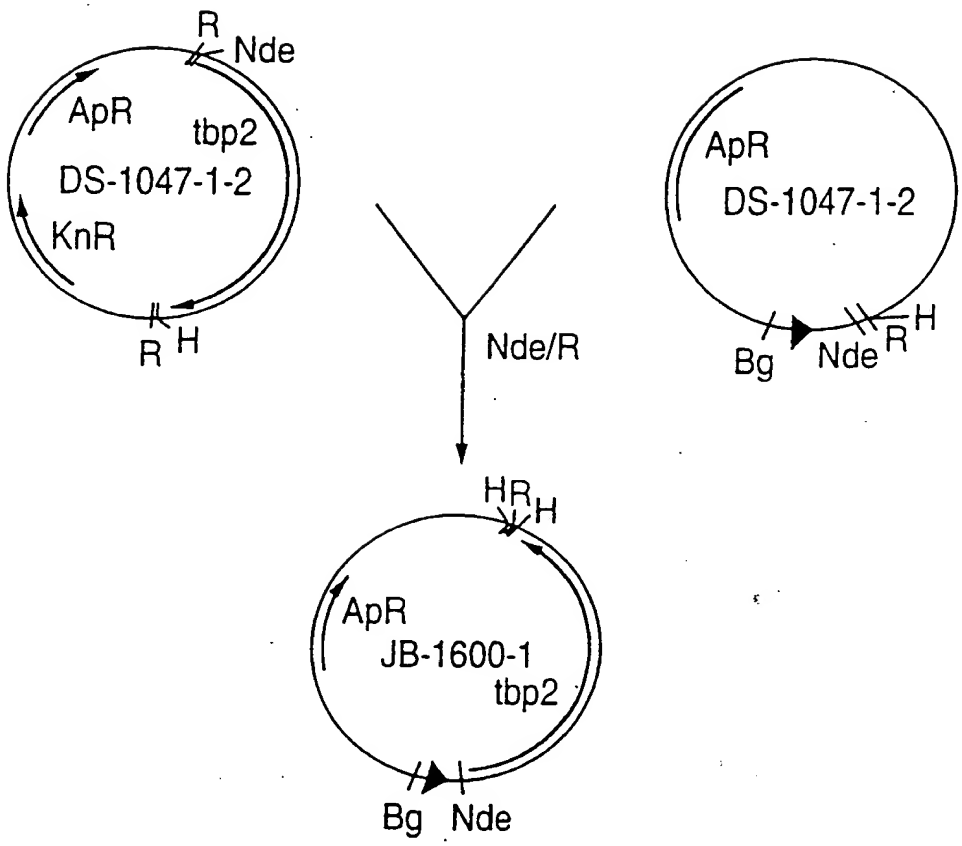

R
Nde
ApR
tbp2
DS-1047-1-2
KnR
R
H
Nde/R
ApR
DS-1047-1-2
Bg
Nde
R
H
HR
H
ApR
JB-1600-1
tbp2
Bg
Nde

FIG.21.

FIG.20A.

Sequence of oligonucleotide pairs (A, B, C and D) for constructing TBP1 and TBP2 expression plasmids

Oligonucleotide pair A (Seq. ID 86 and 87) to join the T7 promoter and Eagan TBP1 gene

Nde I

```
TATGGAAACTCAAAGTATAAAAGATACAAAAGAAGCTATATCATCTGAAGT...
  ACCTTTGAGTTTCATATTTTCTATGTTTTCTTCGATATAGTAGACTTCA...
                                                    Pst I
                          ...GGACACTCAAAGTACAGAAGATTCAGAATTAGAAACTATCTCAGTCACTGCA
                          ...CCTGTGAGTTTCATGTCTTCTAAGTCTTAATCTTTGATAGAGTCAGTG
```

Oligonucleotide pair B (Seq. ID 88 and 89) to join the T7 promoter and Eagan TBP2 genes throught the *E. coli* lpp leader

Nde I

```
TATGAAAGCTACTAAACTGGTTCTGGGTGCTGTTATCCTGGGTTCCACTCTG...
  ACTTTCGATGATTTGACCAAGACCCACGACAATAGGACCCAAGGTGAGAC...
                                                    Ear I
                  ...CTGGCTGGTTGTAGCGGAGGTGGTTGTTTTGATGTAGATAACGTCTCTAATACCCCCTCTTCT
                  ...GACCCACCAACATCGCCTCCACCAACAAAACTACATCTATTGCAGAGATTATGGGGGAGAAGATTT
```

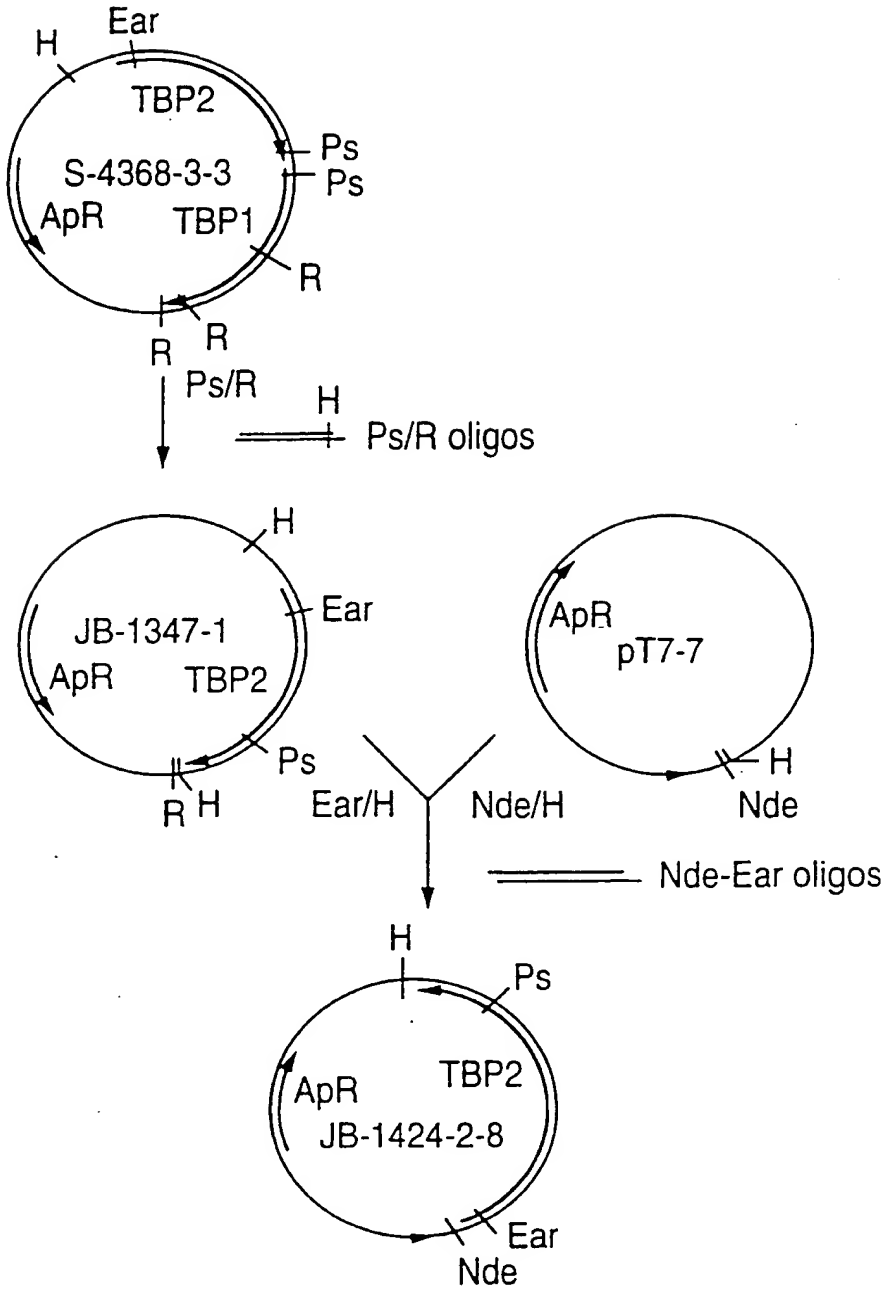

H
Ear
TBP2
Ps
Ps
S-4368-3-3
ApR
TBP1
R
R
R
Ps/R
H
Ps/R oligos
H
Ear
JB-1347-1
ApR
TBP2
Ps
R
H
ApR
pT7-7
H
Nde
Ear/H
Nde/H
Nde-Ear oligos
H
Ps
ApR
TBP2
JB-1424-2-8
Ear
Nde

FIG.18

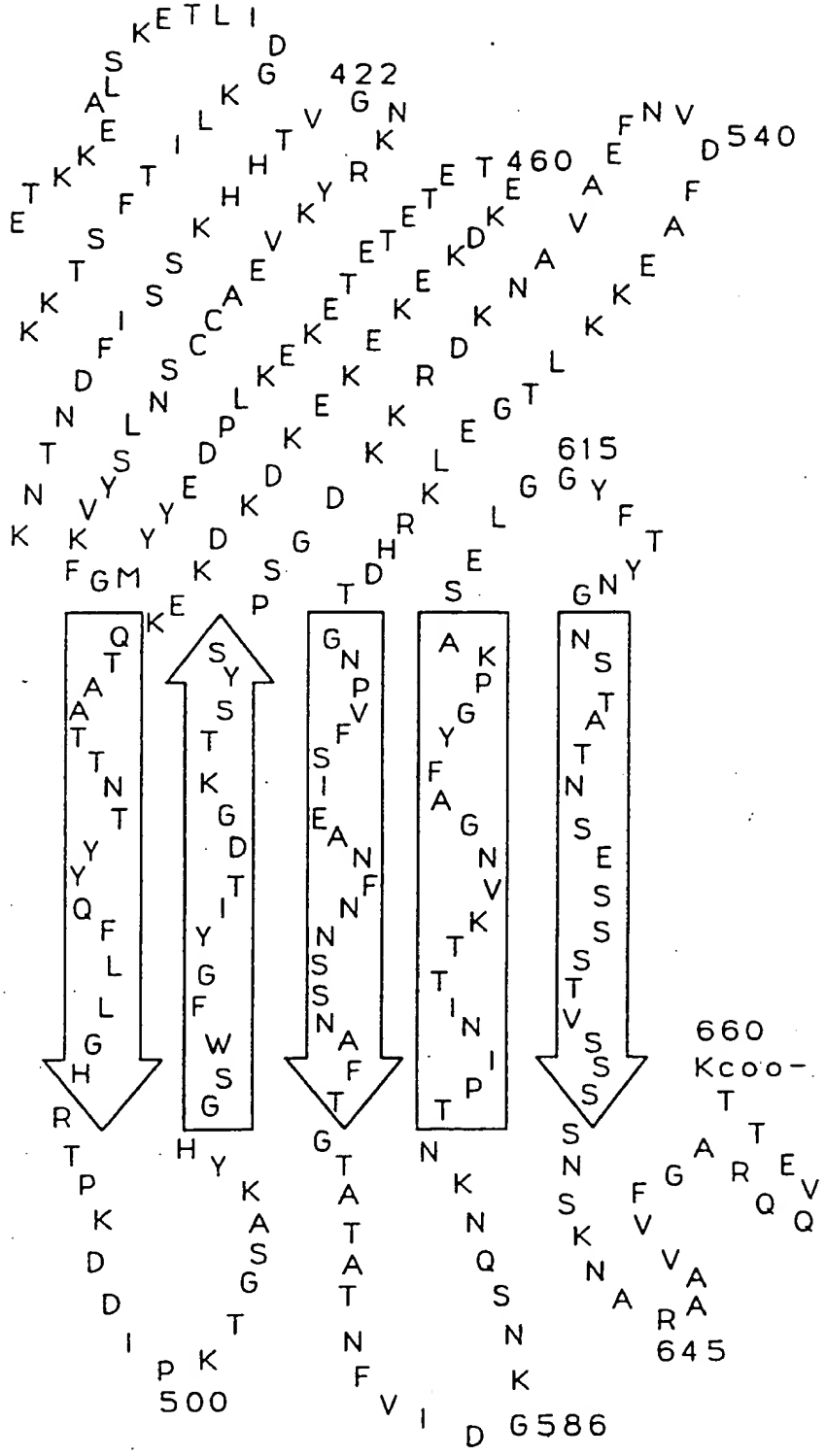


FIG.16B''.

SUBSTITUTE SHEET

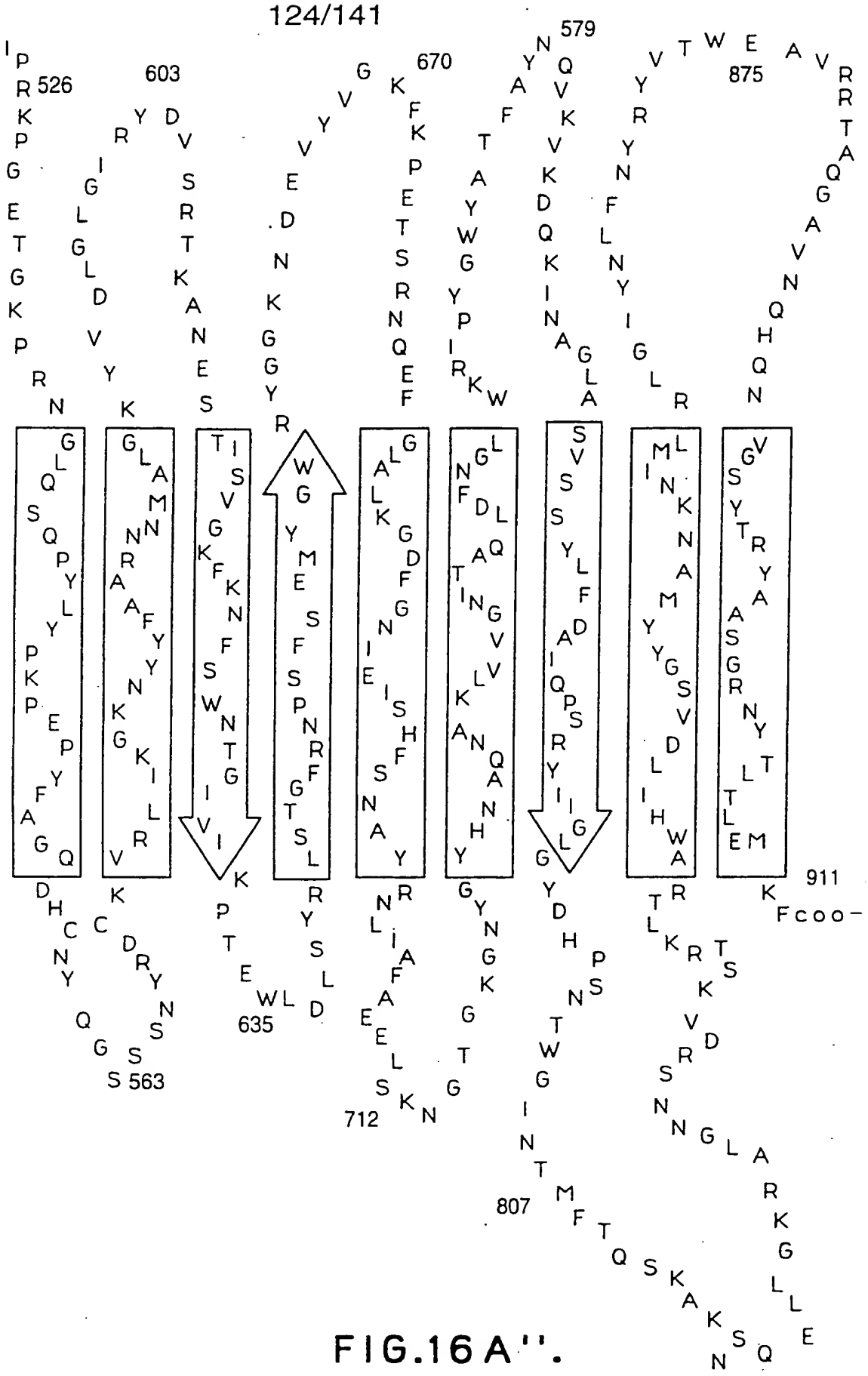


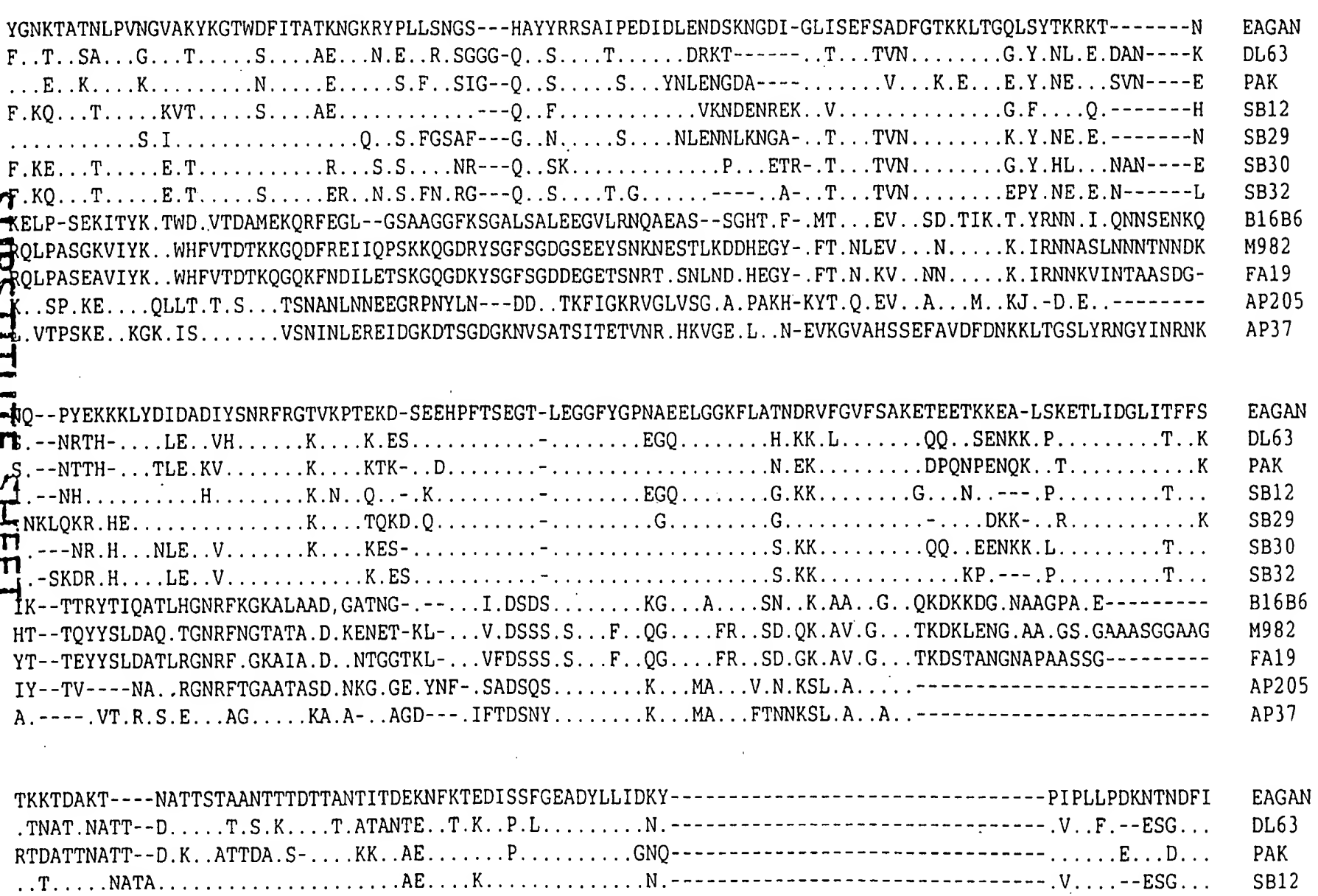
FIG.16A''.

FIG.15D.

```
..AN.-..W.GEASNQEGG.-R...D...ST..IS.T.TAK.RT-S.A.T.T.MIKD--.G.S.V.KTGENG.AL.PQ.TG.SHYTHI-EAT.S.G...KN.    B16B6
H.AN.-..W.GNASD.EGG.-R...T.N..D..I..K.TAENRQ-AQT.T..GMIQG--.G.E...KTAESG.DL.Q..TTRTPKAYITDA..K.G.....    M982
R.AN.-..W.GKASNAT.G.-R.K.T.N.DR.EI..T.TAENRS-EAT.T.D.MIEG--.G.K...KTG.DG.AP.QN..TVTHKVHIANAE.Q.G....N.    FA19
AQVSKENNWVATA.DD.KSGYRT..D...GN.N.S.K.LFDKN.V....TVD.KIDG--.G...K.KTSDEG.AL.SGS.RYE.VKF.DVA-.S.G....T.    AP205
ALVSKG.NWIAEA.NN.ESGYRT..D.N.SD..VN.K.-FDKG.V....TVD.TI.G--.G.I.S.KTSDSG.AL.AGS..HG.AVFSDI-....G....T.    AP37

SELGGYFTYNGN-STATNSESSSTVSSSSNSKNARAAVVFGAR-QQVETT-K*        EAGAN
.........--------NPTDKN.....EK.........KK......-.*        DL63
.........KNP...........P.PP..P..S.......KK-.....N.*        PAK
.........----PTDK.....P.................-.......-.*        SB12
.........---NPTDKN.....P.-..A.........KK-.....N.*          SB29
.........KNP.DK........P.PP..P.........KK-...KNN.*         SB30
.........KDTITK.T......P.PP..P.........KK-.....N.*         SB32
M..S.SFP..APEGKQE---------------K.S.....KR..LVQ*           B16B6
E....W.A.P.DKQ.EKAT-----AT..DGNSASS.T.....KR..PVQ*         M982
E....W.A.P..EQ.KNA------..E.GNGNSASS.T.....KR.KLVK*        FA19
....Q.HHKSENGSVGA--------------------....K-...KK*          AP205
....Q.HHKSDNGSVGA--------------------....K-R.I.K*         AP37
```

FIG.15B:



SUBSTITUTE SHEET

FIG.14C.

```
.............I.....................................................A.........N.SD.............  SB33
G..A.VQD.VR..RWA.V.A.....YRS.HSEDKSV.T.THR.L...A.V.L..FT.M..T..A.....L...A......A.---ESLKTLDL...K.F.  B16B6
G..A.V.D.VR..RWA.V.A.L...YRS.HSDDGSV.T.THRTL...A...L..AD....T..A.....L...A......S.---VQSKAV.ID..K.F.  M982
S..A.V.D.VR..RWA.V.A.L...YRS.HSDDGSV.T.THRTL...A...L..AD....T..A.....L...A......S.---.KIKAV.ID..K.F.  FA19

QFFGLALKGDFGNIEISHFSNAYRNLIAFAEELSKNG-TGKGNY--GYHNAQNAKLVGVNITAQLDFNGLWKRIPYGWYATFAYNQVKVKDQKINAGLAS  EAGAN
.....................................T.....--.....................................R.............  DL63
.................................N...-...A..--.....................................R.............  PAK
.....................................T.....--.....................................R.............  SB33
R.A.IVF.....L.A.Y.N....D....GY.TRTQNGQTSASGDP..R.....RIA.I..LGKI.WH.V.GGL.D.L.S.L...RI....AD.R.DRTF  B16B6
K.A.IVF.....L.A.W.N....D..VRGY.AQIKNGKEEAKGDPA.L...S.RIT.I..LGKI.W..V.DKL.E...S.....R.H.R.I.KR.DRTD  M982
K.A.IVF.....L.A.W.N....D..VRGY.AQIKDGKEQVKGNPA.L...S.RIT.I..LGKI.W..V.DKL.E...S.....R.R.R.I.KR.DRTD  FA19

VSSYLFDAIQPSRYIIGLGYDHPSNTWGINTMFTQSKAKSQNELLGKRALGNNSRD-VKSTRKLTRAWHILDVSGYYMANKNIMLRLGIYNLFNYRYVTW  EAGAN
.....................K.....................N-.....................V.RS.LF...V...L.......  DL63
.....................................................-.....................................  PAK
.....................................Q........N-.....................................  SB33
.....V.....VL.......DGI........Y.....VD....SQ..L.GNANAK.AASRR..P.YVT......NIK.HLT..A.V...L.......  B16B6
.H.........VV.....Q.EGK..V.G.L.Y....EIT....S...L.GNSRNT.A.ARR..P.Y.V......TIK.HFT..A.V...L.......  M982
IQ.H.........VV.S...Q.EGK..V.G.L.Y....EIT....S...L.GNSRNT.A.ARR..P.Y.V......TVK.HFT..A.V...L.H.....  FA19

EAVRQTAQGAVNQHQNVGSYTRYAASGRNYTLTLEMKF*  EAGAN
.................N.................*  DL63
.....................................*  PAK
.....................................*  SB33
.N.....G......K...V.N....P.....FS.....*  B16B6
.N.....G....G.K...V.N....P.....FS.....*  M982
.N.....A......K...V.N....P.....FS.....*  FA19
```

SUBSTITUTE SHEET

FIG.14A.

Comparison of TBP1 amino acid sequences

```
MTKKPYFRLSIISCLLISCYVKAETQSIKDTKEAISSEVDTQSTEDSELETISVTAEKIRDRKDNEVTGLGKIIKTSESISREQVLNIRDLTRYDPGISV   EAGAN
.....V.....   DL63
.....   PAK
.....   SB33
QQQHL...N.L-..SLMTALPVYAENTQAEQAQEKQ----------.D..Q.K.K.QKT.R.....LV.S.DTL.K.....A.   B16B6
QQQHL...N.L-..SLMTALP.YAENTQAGQAQEKQ----------.D..Q.K.K.QKT.R.....LV..ADTL.K....D.....A.   M982
QQQHL...N.L-..SLMTALP.YAENTQAGQAQEKQ----------.D..Q.K.K.QKT.R.....LV..ADTL.K....D.....A.   FA19

VEQGRGASSGYSIRGMDRNRVALLVDGLPQTQSYVVQSPLVARSGYSGTGAINEIEYENVKAVEISKGGSSSEYGNCALAGSVTFQSKSAADILEGDKSW   EAGAN
.....   DL63
.....   PAK
.....   SB33
.....K...S.T...VS.I...TA.AA.GGTRTAGSS.....SN.....A..T.T....IGEG.Q.   B16B6
.....K...S.T....A.I...TA.AA.GGTRTAGSS.....SN.V.Q.S.....A..T.T.D.VIGEGRQ.   M982
.....K...S.T....A.I...TA.AA.GGTRTAGSS.....SN.V.Q.S.....A..T.T.D.VIGEGRQ.   FA19

GIQTKNAYSSKNKGFTHSLAVAGKQGGFEGLAIYTQRNSIETQVHKDALKGVQSYDRLIATTDKSSGYFVIQG-----ECPNGDDK--CAA--KPPATLS   EAGAN
.....V....H.....F....EDQ.A...M.D-----..LD.Y..--.KTSP.R.....   DL63
.....D.V.....H....KPEDQ.A...M.D-----...KP..YNS.LPFA.R..I..   PAK
.....V.....E.....E.F.....-----.....--...--....KV.   SB33
...S.T...G.DHAL.Q...L..RS..A.A.L...K.RGR.IHA....G.....FN..VLDE..KE.GSQYRYFIVEE..H..YAA--.KNKL.ED.SVK   B16B6
...S.T...G..R.L.Q.I.L..RI..A.A.L.H.K.RGG.IRA.E..GR....FN..VLVE.----SSEYAYFIVED..EGKNYET-.KSKP.KDVVGK   M982
...S.T...G..R.L.Q.I.L..RI..A.A.L.H.K.RGG.IRA.EA.GR....FN..APVD.----GSKYAYFIVEE..K..GHEK-.K.NP.KDVVGE   FA19
```

FIG.12B.

GTAGAAACAA CCAAATAATG GAATACTAAA AATGACTAAA AAACCCTATT TTCGCCTAAG

T

GTAGAAACAA CCAAATAATG GAATACTAAA AATGACTAAA AAACCCTATT TTCGCCTAAG

T

GTAGAAACAA CCAAGTAATG GAATACTAAA AATGACTAAA AAACCCTATT TTCGCCTAAG

T

GTAGAAACAA CCAACAAGTA AAAACAACCA AGTAATGGAA TACTAAAAAT GACTAAAAAA

CCCTATTTTC GCCTAAGT

GTAGAAACAA CCAAATAATG GAATACTAAA AATGACTAAA AAA

GTAGAAACAA CCAACAAGTA AAAACAACCA AGTAATGGAA TACTAAAAAT GACTAAAAAA

GTAGAAAAAA ACAACTAGTA AAAACAACCA AGTAATGGAA TACTAAAAAT GACTAAAAAA

GTAGAAACAA CCAACAAGTA GAAAAAAACA AATAATGGAA TACTAAAAAT GACTAAAAAA

TCTAGAAGCT TTTTTAGTCA TTTTTAGTAT TCCAT

FIG.11G.

```
GGA CCT GAT GCT TCT GAA TTA GGC GGT TAT TTC ACC TAT AAC GGA AAA
Gly Pro Asp Ala Ser Glu Leu Gly Gly Tyr Phe Thr Tyr Asn Gly Lys
        580                 585                 590

GAC ACT ATA ACT AAA ACT AAT ACT GAA AGT TCC TCA ACC GTA CCT TCA CCA
Asp Thr Ile Thr Lys Asn Thr Glu Ser Ser Ser Thr Val Pro Ser Pro
    595                 600                 605

CCC AAT TCA CCA AAT GCA AGA GCT GCA GTT GTG TTT GGA GCT AAA AAA
Pro Asn Ser Pro Asn Ala Arg Ala Ala Val Val Phe Gly Ala Lys Lys
610                 615                 620

CAA GTA GAA ACA ACC AAC AAG TAGAAAAAA CAAATAATGG AATACTAAAA
Gln Val Glu Thr Thr Asn Lys
625                 630

ATGACTAAAA AAGCTTCTAG AAAGCCGAAT TC
```

FIG.11E.

```
GAT AAT TAC CCT ATT CCG CTT TTA CCT GAG AGT GGT GAT TTC ATA AGT
Asp Asn Tyr Pro Ile Pro Leu Leu Pro Glu Ser Gly Asp Phe Ile Ser
385                 390                 395                 400

AGT AAG CAC CAT GAG GTA GGA GGT AAA CGC TAT AAA GTG GAA GCA TGT
Ser Lys His His Glu Val Gly Gly Lys Arg Tyr Lys Val Glu Ala Cys
                405                 410                 415

TGC AAG AAT CTA TGC TAT GTG AAA TTT GGT ATG TAT TAT GAG GAT AAA
Cys Lys Asn Leu Cys Tyr Val Lys Phe Gly Met Tyr Tyr Glu Asp Lys
            420                 425                 430

GAG AAC AAC AAA AAT GAA ACA GAC AAA GAA AAA GAA AAA CAA ACG ACA
Glu Asn Asn Lys Asn Glu Thr Asp Lys Glu Lys Glu Lys Gln Thr Thr
        435                 440                 445

ACA TCT ATC AAG ACT TAT TAT CAA TTC TTA TTA GGT CTC CGG ACT CCC
Thr Ser Ile Lys Thr Tyr Tyr Gln Phe Leu Leu Gly Leu Arg Thr Pro
    450                 455                 460

AGT TCT GAA ATT CCT AAA ATG GGA AAC GTG ACA TAT CGC GGT AGT TGG
Ser Ser Glu Ile Pro Lys Met Gly Asn Val Thr Tyr Arg Gly Ser Trp
465                 470                 475                 480
```

FIG.11C.

```
GGC AAA AAT TAT TCT TTG TTC AAT AAT AGA GGT CAA GCT TAT TCT CGA
Gly Lys Asn Tyr Ser Leu Phe Asn Asn Arg Gly Gln Ala Tyr Ser Arg
        195                 200                 205

CGT AGT GCT ACT CCA GGA GAT ATT GAT TTA GAA AAC GGT GAC GCA GGC
Arg Ser Ala Thr Pro Gly Asp Ile Asp Leu Glu Asn Gly Asp Ala Gly
    210                 215                 220

TTA ACA AGT GAA TTT ACT GTC AAT TTT GGT ACA AAA AAG CTC ACT GGA
Leu Thr Ser Glu Phe Thr Val Asn Phe Gly Thr Lys Lys Leu Thr Gly
225                 230                 235                 240

GAA CCT TAT TAT AAT GAA AGG GAA ACA AAT CTT AAT CAA TCA AAA GAT
Glu Pro Tyr Tyr Asn Glu Arg Glu Thr Asn Leu Asn Gln Ser Lys Asp
                245                 250                 255

AGA AAA CAT AAA CTC TAC GAT CTA GAA GCT GAT GTG TAT AGC AAC CGA
Arg Lys His Lys Leu Tyr Asp Leu Glu Ala Asp Val Tyr Ser Asn Arg
            260                 265                 270

TTC AGA GGT ACA GTA AAG CCA ACC AAA AAA GAG TCT TCT GAA GAA CAT
Phe Arg Gly Thr Val Lys Pro Thr Lys Lys Glu Ser Ser Glu Glu His
        275                 280                 285
```

FIG.11A.

```
ATG AAA TCT GTA CCT CTT ATC TCT GGT GGA CTT TCC TTT TTA CTA AGT
Met Lys Ser Val Pro Leu Ile Ser Gly Gly Leu Ser Phe Leu Leu Ser
  1               5                  10                  15

GCT TGT AGC GGA GGG GGG TCT TTT GAT GTA GAT AAC GTC TCT AAT ACC
Ala Cys Ser Gly Gly Gly Ser Phe Asp Val Asp Asn Val Ser Asn Thr
             20                  25                  30

CCC TCT TCT AAA CCA CGT TAT CAA GAC GAT ACC TCG AAT CAA AGA ACA
Pro Ser Ser Lys Pro Arg Tyr Gln Asp Asp Thr Ser Asn Gln Arg Thr
         35                  40                  45

AAA TCT AAA TTG GAA AAG TTG TCC ATT CCT TCT TTA GGA GGA GGG ATG
Lys Ser Lys Leu Glu Lys Leu Ser Ile Pro Ser Leu Gly Gly Gly Met
     50                  55                  60

AAG TTA GTT GTG CAA AAT TTT GCT GGT GCT AAA GAA CCT AGT TTC TTA
Lys Leu Val Val Gln Asn Phe Ala Gly Ala Lys Glu Pro Ser Phe Leu
 65                  70                  75                  80

AAT GAA AAT GAC TAT ATA TCA TAT TTT TCC TCA CTT TCT ATG ATT AAA
Asn Glu Asn Asp Tyr Ile Ser Tyr Phe Ser Ser Leu Ser Met Ile Lys
                 85                  90                  95
```

FIG.10F.

```
AGT TGG TTT GGT TAT ATT GGT GAT GAC AAG ACA TCT TAC TCC ACT ACT
Ser Trp Phe Gly Tyr Ile Gly Asp Asp Lys Thr Ser Tyr Ser Thr Thr
                485                 490                 495

GGA GAT AAA AAT GCT CTC GCC GAG TTT GAT GTA AAT TTT ACC GAT AAA
Gly Asp Lys Asn Ala Leu Ala Glu Phe Asp Val Asn Phe Thr Asp Lys
            500                 505                 510

AAG CTA ACA GGC GAA TTA AAA CGA GCC GAT AAT CAA AAT ACC GTA TTT
Lys Leu Thr Gly Glu Leu Lys Arg Ala Asp Asn Gln Asn Thr Val Phe
        515                 520                 525

AGA ATT AAT GCA GAC TTT AAA AAT AAT GAT AAT GCC TTC AAA GGT ACA
Arg Ile Asn Ala Asp Phe Lys Asn Asn Asp Asn Ala Phe Lys Gly Thr
    530                 535                 540

GCA ACC GCA GAA AAT TTT GTA ATA GAT GGT AAC AAT AGT CAA ACT GGA
Ala Thr Ala Glu Asn Phe Val Ile Asp Gly Asn Asn Ser Gln Thr Gly
545                 550                 555                 560

AAT ACC CAA ATT AAT ATT AAA ACT GAA GTA AAT GGG GCA TTT TAT GGT
Asn Thr Gln Ile Asn Ile Lys Thr Glu Val Asn Gly Ala Phe Tyr Gly
                565                 570                 575
```

FIG.10D.

```
CCC TTT ACC AGC GAG GGA ACA TTA GAA GGT GGT TTT TAT GGG CCT AAT
Pro Phe Thr Ser Glu Gly Thr Leu Glu Gly Gly Phe Tyr Gly Pro Asn
        290                 295                 300

GCT GAA GAA CTA GGG GGA AAA TTT TTA GCT AGC GAT AAA AAA GTT TTT
Ala Glu Glu Leu Gly Gly Lys Phe Leu Ala Ser Asp Lys Lys Val Phe
305                 310                 315                 320

GGG GTA TTT AGT GCC AAA GAA CAG CAA GAA ACG GAA GAA AAC AAA AAA
Gly Val Phe Ser Ala Lys Glu Gln Gln Glu Thr Glu Glu Asn Lys Lys
                325                 330                 335

TTA CTC AAA GAA ACC TTA ATT GAT GGC AAG CTA ACT ACT ACT TTC TCT ACT
Leu Leu Lys Glu Thr Leu Ile Asp Gly Lys Leu Thr Thr Phe Ser Thr
        340                 345                 350

AAA AAA ACC AAT GCA ACA ACC GAT GCA ACA ACC AGT ACA ACA ACC AGT
Lys Lys Thr Asn Ala Thr Thr Asp Ala Thr Thr Ser Thr Thr Thr Ser
        355                 360                 365

ACA GCA ACC AAT GCA ACA GCC GAT GCA GAA AAC TTT ACG ACA AAA GAT
Thr Ala Thr Asn Ala Thr Ala Asp Ala Glu Asn Phe Thr Thr Lys Asp
        370                 375                 380
```

FIG.10B.

```
GAT GAT GTT GAA AAT AAC AAT ACA AAC GGG GGG GAC TAT ATT GGC TCA
Asp Asp Val Glu Asn Asn Asn Thr Asn Gly Gly Asp Tyr Ile Gly Ser
  100                 105                 110

ATA GAC GAG CCT AGT ACA ACA AAT CCA CTC GAA AAG CAT CAT GGA CAA
Ile Asp Glu Pro Ser Thr Thr Asn Pro Leu Glu Lys His His Gly Gln
  115                 120                 125

AGG TAT GTA TAT TCA GGG CTT TAT TAT ATT CAA TCG TGG AGT CTA AGA
Arg Tyr Val Tyr Ser Gly Leu Tyr Tyr Ile Gln Ser Trp Ser Leu Arg
  130                 135                 140

GAT TTA CCA AAG AAG TTT TAT TCA GGT TAC TAT GGA TAT GCG TAT TAC
Asp Leu Pro Lys Lys Phe Tyr Ser Gly Tyr Tyr Gly Tyr Ala Tyr Tyr
145                 150                 155                 160

TTT GGC AAG GAA ACA GCC ACT ACA TTA CCT GTA AAT GGC GAA GCA ACG
Phe Gly Lys Glu Thr Ala Thr Thr Leu Pro Val Asn Gly Glu Ala Thr
                165                 170                 175

TAT AAA GGA ACT TGG GAT TTC ATC ACT GCA ACT AGA AAT GGC AAA AGT
Tyr Lys Gly Thr Trp Asp Phe Ile Thr Ala Thr Arg Asn Gly Lys Ser
                180                 185                 190
```

FIG.9G.

```
ACG GCA ACA GTA AAC GGG GCA TTT TAT GGA CCT AAG GCT ACA GAA TTA
Thr Ala Thr Val Asn Gly Ala Phe Tyr Gly Pro Lys Ala Thr Glu Leu
        575                 580                 585

GGC GGT TAT TTC ACT TAT AAC GGA AAC AAT CCT ACA GAT AAA AAT TCC
Gly Gly Tyr Phe Thr Tyr Asn Gly Asn Asn Pro Thr Asp Lys Asn Ser
        590                 595                 600

TCA ACC GTT TCA CCA TCC AAT TCA GCA AAT GCT CGT GCT GCC GTT GTG
Ser Thr Val Ser Pro Ser Asn Ser Ala Asn Ala Arg Ala Ala Val Val
        605                 610                 615

TTT GGC GCT AAA AAA CAA GTA GAA ACA ACC AAC AAG TAAAACAAC
Phe Gly Ala Lys Lys Gln Val Glu Thr Thr Asn Lys
620                 625                 630

CAAGTAATGG AATACTAAAA ATGACTAAAA AAGCTTCTAG AAAGCCGAAT TC
```

FIG.9E.

```
GAT TAC CTT TTA ATT GAT AAT TAC CCT GTT CCT CTT TTC CCT GAA GAA
Asp Tyr Leu Leu Ile Asp Asn Tyr Pro Val Pro Leu Phe Pro Glu Glu
380                 385                 390                 395

AAT ACT AAT GAT TTC ATA ACT AGT AGG CAC CAi' AAG GTA GGA GAT AAA
Asn Thr Asn Asp Phe Ile Thr Ser Arg His His Lys Val Gly Asp Lys
                400                 405                 410

ACC TAT AAA GTA GAA GCA TGT TGC AAG AAT CTA AGC TAT GTG AAA TTT
Thr Tyr Lys Val Glu Ala Cys Cys Lys Asn Leu Ser Tyr Val Lys Phe
            415                 420                 425

GGT ATG TAT TAT GAA GAC CCA TTA AAT GGA GAA AAT GGC AAA GAA AAA
Gly Met Tyr Tyr Glu Asp Pro Leu Asn Gly Glu Asn Gly Lys Glu Lys
        430                 435                 440

GAA AAA GAA AAA GAA AAA GAC AAA GAA AAA CAA GCG ACA ACA TCT ATC
Glu Lys Glu Lys Glu Lys Asp Lys Glu Lys Gln Ala Thr Thr Ser Ile
    445                 450                 455

AAG ACT TAT TAT CAA TTC TTA TTA GGT CAC CGT ACT GCC AAG GCC GAC
Lys Thr Tyr Tyr Gln Phe Leu Leu Gly His Arg Thr Ala Lys Ala Asp
460                 465                 470                 475
```

FIG.9C.

```
AAT GGC CAA CGT TAT TCT TTA TTT GGT AGC GCT TTT GGA GCT TAT AAT
Asn Gly Gln Arg Tyr Ser Leu Phe Gly Ser Ala Phe Gly Ala Tyr Asn
                190                 195                 200

AGA CGC AGT GCT ATT TCA GAA GAT ATA GAT AAT TTA GAA AAT AAT CTA
Arg Arg Ser Ala Ile Ser Glu Asp Ile Asp Asn Leu Glu Asn Asn Leu
            205                 210                 215

AAG AAT GGT GCG GGA TTA ACT AGT GAA TTT ACT GTC AAT TTT GGT ACG
Lys Asn Gly Ala Gly Leu Thr Ser Glu Phe Thr Val Asn Phe Gly Thr
220                 225                 230                 235

AAA AAG CTC ACT GGA AAA CTT TAT TAT AAT GAA AGG GAA ACA AAT CTT
Lys Lys Leu Thr Gly Lys Leu Tyr Tyr Asn Glu Arg Glu Thr Asn Leu
                240                 245                 250

AAT AAA TTA CAA AAG AGA AAA CAT GAA CTC TAT GAT ATA GAT GCC GAT
Asn Lys Leu Gln Lys Arg Lys His Glu Leu Tyr Asp Ile Asp Ala Asp
            255                 260                 265

ATT TAT AGT AAT AGA TTC AGA GGT AAA GTA AAG CCA ACA ACC CAA AAA
Ile Tyr Ser Asn Arg Phe Arg Gly Lys Val Lys Pro Thr Thr Gln Lys
            270                 275                 280
```

FIG.9A.

```
GAATTCGGCT TGGATCCAT ATG AAA TCT GTA CCT CTT ATC TCT GGT GGA CTT
                Met Lys Ser Val Pro Leu Ile Ser Gly Gly Leu
                  1               5                  10

TCC TTT TTA CTA AGT GCT TGT AGC GGA GGG GGG TCT TTT GAT GTA GAT
Ser Phe Leu Leu Ser Ala Cys Ser Gly Gly Gly Ser Phe Asp Val Asp
            15                  20                  25

AAC GTC TCT AAT CCA TCC TCT TCT AAA CCA CGT TAT CAA GAC GAT ACT
Asn Val Ser Asn Pro Ser Ser Ser Lys Pro Arg Tyr Gln Asp Asp Thr
        30                  35                  40

TCA AGT TCA AGA ACA AAA TCT AAT TTG AAA AAG TTG TCC ATT CCT TCT
Ser Ser Ser Arg Thr Lys Ser Asn Leu Lys Lys Leu Ser Ile Pro Ser
    45                  50                  55

TTA GGG GGA GGG ATG AAG TTA GTG GCT CAG AAT CTT AGT GAT AAG AAC
Leu Gly Gly Gly Met Lys Leu Val Ala Gln Asn Leu Ser Asp Lys Asn
60                  65                  70                  75

AAA CCT AGT CTC TTA AAT GAA GAT GAC TAT ATA TCA TAT TTT TCC TCA
Lys Pro Ser Leu Leu Asn Glu Asp Asp Tyr Ile Ser Tyr Phe Ser Ser
                80                  85                  90
```

FIG.8E.

```
GAA AAC TTT AAG ACG AAA GAT ATA TCA AGT TTT GGT GAA GCT GAT TAC
Glu Asn Phe Lys Thr Lys Asp Ile Ser Ser Phe Gly Glu Ala Asp Tyr
385                 390                 395

CTT TTA ATT GAT AAT TAC CCT GTT CCT CTT TTA CCT GAG AGT GGT GAT
Leu Leu Ile Asp Asn Tyr Pro Val Pro Leu Leu Pro Glu Ser Gly Asp
400                 405                 410                 415

TTC ATA AGT AGT AAG CAC CAT ACT GTA GGA AAG AAA ACC TAT CAA GTA
Phe Ile Ser Ser Lys His His Thr Val Gly Lys Lys Thr Tyr Gln Val
                420                 425                 430

AAA GCA TGT TGC AGT AAT CTA AGC TAT GTG AAA TTT GGT ATG TAT TAT
Lys Ala Cys Cys Ser Asn Leu Ser Tyr Val Lys Phe Gly Met Tyr Tyr
            435                 440                 445

GAA GTC CCA CCT AAA GAA GAA AAA GAC AAA GAA AAA AAA GAA AAA AAA
Glu Val Pro Pro Lys Glu Glu Glu Lys Asp Lys Glu Lys Lys Glu Lys
        450                 455                 460

GAA AAA GAA AAA CAA GCG ACA AAT CTA TCG AAC ACT TAT TAT CAA TTC
Glu Lys Glu Lys Gln Ala Thr Asn Leu Ser Asn Thr Tyr Tyr Gln Phe
    465                 470                 475
```

FIG.8C.

```
GGC AAA AGG TAT CCT TTG TTA AGT AAT GGC AGT CAA GCT TAT TTT CGA
Gly Lys Arg Tyr Pro Leu Leu Ser Asn Gly Ser Gln Ala Tyr Phe Arg
                195                 200                 205

CGT AGT GCA ATT CCA GAA GAT ATT GAT TTA GAA GTT AAA AAT GAT GAG
Arg Ser Ala Ile Pro Glu Asp Ile Asp Leu Glu Val Lys Asn Asp Glu
            210                 215                 220

AAT AGA GAA AAA GGG CTA GTG AGT GAA TTT AGT GCA GAT TTT GGG ACT
Asn Arg Glu Lys Gly Leu Val Ser Glu Phe Ser Ala Asp Phe Gly Thr
        225                 230                 235

AAA AAA CTG ACA GGA GGA CTG TTT TAC ACC AAA AGA CAA ACT CAT ATT
Lys Lys Leu Thr Gly Gly Leu Phe Tyr Thr Lys Arg Gln Thr His Ile
    240                 245                 250                 255

CAA AAC CAT GAA AAG AAA AAA CTC TAT GAT ATA GAT GCC CAT ATT TAT
Gln Asn His Glu Lys Lys Lys Leu Tyr Asp Ile Asp Ala His Ile Tyr
                260                 265                 270

AGT AAT AGA TTC AGA GGT AAA GTA AAT CCT ACC CAA AAA GAT TCT AAA
Ser Asn Arg Phe Arg Gly Lys Val Asn Pro Thr Gln Lys Asp Ser Lys
            275                 280                 285
```

FIG.8A.

```
AT ATG AAA TCT GTA CCT CTT ATC TCT GGT GGA CTT TCC TTT TTA TTA
   Met Lys Ser Val Pro Leu Ile Ser Gly Gly Leu Ser Phe Leu Leu
     1               5                  10                  15

AGT GCT TGT AGC GGG GGA GGT GGT TCT TTT GAT GTA GAT GAC GTC TCT
Ser Ala Cys Ser Gly Gly Gly Gly Ser Phe Asp Val Asp Asp Val Ser
                20                  25                  30

AAT CCC TCC TCT TCT AAA CCA CGT TAT CAA GAC GAT ACT TCA AGT TCA
Asn Pro Ser Ser Ser Lys Pro Arg Tyr Gln Asp Asp Thr Ser Ser Ser
            35                  40                  45

AGA ACA AAA TCT AAA TTG GAA AAT TTG TCC ATT CCT TCT TTA GGG GGA
Arg Thr Lys Ser Lys Leu Glu Asn Leu Ser Ile Pro Ser Leu Gly Gly
        50                  55                  60

GGG ATG AAG TTA GTG GCT CAG AAT CTT CGT GAT AGG ACA AAA CCT AGT
Gly Met Lys Leu Val Ala Gln Asn Leu Arg Asp Arg Thr Lys Pro Ser
    65                  70                  75

CTC TTA AAT GAA GAT GAC TAT ATG ATA TTT TCC TCA CTT TCA ACG ATT
Leu Leu Asn Glu Asp Asp Tyr Met Ile Phe Ser Ser Leu Ser Thr Ile
80                  85                  90                  95
```

FIG.7M.

```
TAT AAC CGA GTA AAA GTT AAA GAT CAA AAA ATC AAT GCT GGT TTG GCC
Tyr Asn Arg Val Lys Val Lys Asp Gln Lys Ile Asn Ala Gly Leu Ala
                930                 935                 940

TCC GTA AGC AGT TAT TTA TTT GAT GCC ATT CAG CCC AGC CGT TAT ATC
Ser Val Ser Ser Tyr Leu Phe Asp Ala Ile Gln Pro Ser Arg Tyr Ile
            945                 950                 955

ATT GGT TTA GGC TAT GAT CAT CCA AGT AAT ACT TGG GGA ATT AAT ACA
Ile Gly Leu Gly Tyr Asp His Pro Ser Asn Thr Trp Gly Ile Asn Thr
        960                 965                 970

ATG TTT ACT CAA TCA AAA GCA AAA TCT CAA AAT GAA TTG CTA GGA CAA
Met Phe Thr Gln Ser Lys Ala Lys Ser Gln Asn Glu Leu Leu Gly Gln
    975                 980                 985

CGT GCA TTG GGT AAC AAT TCA AGG AAT GTA AAA TCA ACA AGA AAA CTT
Arg Ala Leu Gly Asn Asn Ser Arg Asn Val Lys Ser Thr Arg Lys Leu
990                 995                 1000                1005

ACT CGG GCA TGG CAT ATC TTA GAT GTA TCG GGT TAT TAC ATG GCG AAT
Thr Arg Ala Trp His Ile Leu Asp Val Ser Gly Tyr Tyr Met Ala Asn
                1010                1015                1020
```

FIG. 7K.

```
AGC GAC TGT AAA GTG CGG TTA ATT AAA GGG AAA AAT TAT TAT TTC GCA
Ser Asp Cys Lys Val Arg Leu Ile Lys Gly Lys Asn Tyr Tyr Phe Ala
  735                 740                 745

GCA CGC AAT AAT ATG GCA TTA GGG AAA TAC ATT GAT TTA GGT TTA GGT
Ala Arg Asn Asn Met Ala Leu Gly Lys Tyr Ile Asp Leu Gly Leu Gly
  750                 755                 760                 765

ATT CGG TAT GAC GTA TCT CGT ACA AAA GCT AAT GAA TCA ACT ATT AGT
Ile Arg Tyr Asp Val Ser Arg Thr Lys Ala Asn Glu Ser Thr Ile Ser
                770                 775                 780

GTT GGT AAA TTT AAA AAT TTC TCT TGG AAT ACT GGT ATT GTC ATA AAA
Val Gly Lys Phe Lys Asn Phe Ser Trp Asn Thr Gly Ile Val Ile Lys
                785                 790                 795

CCA ACG GAA TGG CTT GAT CTT TCT TAT CGC CTT TCT ACT GGA TTT AGA
Pro Thr Glu Trp Leu Asp Leu Ser Tyr Arg Leu Ser Thr Gly Phe Arg
  800                 805                 810

AAT CCT AGT TTT GCT GAA ATG TAT GGT TGG CGG TAT GGT GGC AAT AAT
Asn Pro Ser Phe Ala Glu Met Tyr Gly Trp Arg Tyr Gly Gly Asn Asn
  815                 820                 825
```

FIG.71.

```
GAT TAT GGT GCA TAT CAA CGT ATT GAG GAT GGC CGA GGC GTT AAC TAT
Asp Tyr Gly Ala Tyr Gln Arg Ile Glu Asp Gly Arg Gly Val Asn Tyr
545                 550                 555

GCA AGT GGG CTT TAT TTC GAT GAA CAC CAT AGA AAA CAG CGT GTA GGT
Ala Ser Gly Leu Tyr Phe Asp Glu His His Arg Lys Gln Arg Val Gly
560                 565                 570

ATT GAA TAT ATT TAC GAA AAT AAG AAC AAA GCG GGC ATC ATT GAC AAA
Ile Glu Tyr Ile Tyr Glu Asn Lys Asn Lys Ala Gly Ile Ile Asp Lys
575                 580                 585

GCA GTG TTA AGT GCT AAT CAA CAA AAC ATC ATA CTT GAC AGT TAT ATG
Ala Val Leu Ser Ala Asn Gln Gln Asn Ile Ile Leu Asp Ser Tyr Met
590                 595                 600                 605

CGA CAT ACG CAT TGC AGT CTT TAT CCT AAT CCA AGT AAG AAT TGC CGC
Arg His Thr His Cys Ser Leu Tyr Pro Asn Pro Ser Lys Asn Cys Arg
                610                 615                 620

CCG ACA CTT GAT AAA CCT TAT TCA TAC TAT CGT TCT GAT AGA AAT GTT
Pro Thr Leu Asp Lys Pro Tyr Ser Tyr Tyr Arg Ser Asp Arg Asn Val
625                 630                 635
```

FIG. 7G.

```
TCT GTA ACA TTT CAA AGC AAA TCC GCA GCC GAT ATC TTA GAA GGA GAC

Ser Val Thr Phe Gln Ser Lys Ser Ala Ala Asp Ile Leu Glu Gly Asp
350                 355                 360                 365

AAA TCA TGG GGA ATT CAA ACT AAA AAT GCT TAT TCA AGC AAA AAT AAA
Lys Ser Trp Gly Ile Gln Thr Lys Asn Ala Tyr Ser Ser Lys Asn Lys
                370                 375                 380

GGC TTT ACC CAT TCT TTA GCT GTA GCA GGA AAA CAA GGT GGA TTT GAA
Gly Phe Thr His Ser Leu Ala Val Ala Gly Lys Gln Gly Gly Phe Glu
            385                 390                 395

GGG GTC GCC ATT TAC ACT CAA CGA AAT TCG GAG GAA ACC CAA GTC CAT
Gly Val Ala Ile Tyr Thr Gln Arg Asn Ser Glu Glu Thr Gln Val His
        400                 405                 410

AAA GAT GCA TTA AAA GGC GTA CAA AGT TAT GAG CGA TTC ATC GCC ACA
Lys Asp Ala Leu Lys Gly Val Gln Ser Tyr Glu Arg Phe Ile Ala Thr
    415                 420                 425

ACA GAT AAA TCT TCA GGA TAC TTT GTG ATA CAA GGT GAG TGT CCA AAT
Thr Asp Lys Ser Ser Gly Tyr Phe Val Ile Gln Gly Glu Cys Pro Asn
430                 435                 440                 445
```

FIG.7E.

```
TCCAATTCAG CAAATGCTCG TGCTGCCGTT GTGTTTGGAG CTAAAAACA AGTAGACACA

ACCAACAAGT AGAAAAACC AAATAATGGA ATACTAAAA ATG ACT AAA AAA CCC
                                    Met Thr Lys Lys Pro
                                    170

TAT TTT CGC CTA AGT ATT ATT TCT TGT CTT TTA ATT TCA TGC TAT GTA
Tyr Phe Arg Leu Ser Ile Ile Ser Cys Leu Leu Ile Ser Cys Tyr Val
    175                 180                 185

AAA GCA GAA ACT CAA AGT ATA AAA GAT ACA AAA GAA GCT ATA TCA TCT
Lys Ala Glu Thr Gln Ser Ile Lys Asp Thr Lys Glu Ala Ile Ser Ser
190                 195                 200                 205

GAA GTG GAC ACT CAA AGT ACA GAA GAT TCA GAA TTA GAA ACT ATC TCA
Glu Val Asp Thr Gln Ser Thr Glu Asp Ser Glu Leu Glu Thr Ile Ser
                210                 215                 220

GTC ACT GCA GAA AAA ATA AGA GAT CGT AAA GAT AAT GAA GTA ACT GGA
Val Thr Ala Glu Lys Ile Arg Asp Arg Lys Asp Asn Glu Val Thr Gly
            225                 230                 235

CTT GGC AAA ATT ATA AAA ACG AGT GAA AGT ATC AGC CGA GAA CAA GTA
Leu Gly Lys Ile Ile Lys Thr Ser Glu Ser Ile Ser Arg Glu Gln Val
        240                 245                 250
```

FIG.7C.

ACT TTG GCA AGC AAA CAG CCA CTA CAT TAC CTG TAGATGGCGA AGCAACGTAT
Thr Leu Ala Ser Lys Gln Pro Leu His Tyr Leu
160 165

AAAGGAACTT GGCACTTCAT CACCGCAACT GAAAATGGCA AAAAGTATTC TTTGTTCAGT

AATGATAGCG GTCAAGCTTA TCGCAGACGT AGTGCAATTC CAGAAGATAT TGATTTAGAA

AAAAATGATT CAACTAATGG TGACAAGGGC TTAATAAGTG AATTTAGTGT CAATTTTGGT

ACAAAAAAGC TCACTGGAAA ACTTTATTAT AATGAAAGAG AAACAGAACT TAATAAATCA

AAAGATAGAA AACATACACT CTACAATCTA GAAGCTGAAG TGTATAGTAA CCGATTCAGG

GGTACAGTAA AGCCAACCGA AAAAGATTCT ACAGATCATC CCTTTACCAG CGAGGGAACA

TTAGAAGGTG GTTTTTATGG GCCTAAAGGT GAAGAACTAG GAGGAAAGTT TTTAGCTGGC

GATAAAAAAG TTTTTGGGGT ATTTAGTGCC AAAGAAACGG AAGAAACAAA AAAGAAAGCG

TTATCCAAGG AAACCTTAAT TGATGGCAAG CTAACTACTT TTAAAACAAC CAATGCAACA

ACCAATGCAA CAGCCAATGC AACAACCAGT ACAACAGCCA GTACAACAAC CGATGCAGAA

FIG. 7 A.

```
CAACATCTGC CCAAGCTATA TTCGTTAATG ATAAGCCTAT TAATGATAAG CCTATTAATG

ATAAGAAAGA AATTTGTTTT ACGCCATTTT TCATATTTTA TCCATGAACT TAAAAAATTC

TAAGTTGACA TTATTACAAA AAAAGAACAA TAATGCGAAT TATTATCAAT TTTGTATAAG

AATATAATTC T ATG AAA TCT GTA CCT CTT ATC TCT GGT GGA CTT TCC TTT
          Met Lys Ser Val Pro Leu Ile Ser Gly Gly Leu Ser Phe
            1               5                  10

TTA TTA AGT GCT TGT AGC GGA GGA GGG TCT TTT GAT GTA GAT AAC GTC
Leu Leu Ser Ala Cys Ser Gly Gly Gly Ser Phe Asp Val Asp Asn Val
        15                  20                  25

TCT AAT CCC TCC TCT TCT AAA CCA CGT TAT CAA GAC GAT ACC TCG AAT
Ser Asn Pro Ser Ser Ser Lys Pro Arg Tyr Gln Asp Asp Thr Ser Asn
    30                  35                  40                  45

CAA AGA ACA AAA TCT GAT TTG CAA AAG TTG TCC ATT CCT TCT TTA GGG
Gln Arg Thr Lys Ser Asp Leu Gln Lys Leu Ser Ile Pro Ser Leu Gly
                50                  55                  60
```

SUBSTITUTE SHEET

FIG.6P.

```
CGT ATT CCC TAC GGT TGG TAT GCA ACA TTT GCT TAT AAC CGA GTA AAA
Arg Ile Pro Tyr Gly Trp Tyr Ala Thr Phe Ala Tyr Asn Arg Val Lys
      1405                1410                1415

GTT AAA GAT CAA AAA ATC AAT GCT GGT TTG GCC TCC GTA AGC AGT TAT
Val Lys Asp Gln Lys Ile Asn Ala Gly Leu Ala Ser Val Ser Ser Tyr
 1420                1425                1430                1435

TTA TTT GAT GCC ATT CAG CCC AGC CGT TAT ATC ATT GGT TTA GGC TAT
Leu Phe Asp Ala Ile Gln Pro Ser Arg Tyr Ile Ile Gly Leu Gly Tyr
                1440                1445                1450

GAT CAT CCA AGT AAT ACT TGG GGA ATT AAT ACA ATG TTT ACT CAA TCA
Asp His Pro Ser Asn Thr Trp Gly Ile Asn Thr Met Phe Thr Gln Ser
                1455                1460                1465

AAA GCA AAA TCT CAA AAT GAA TTG CTA GGA AAA CGT GCA TTG GGT AAC
Lys Ala Lys Ser Gln Asn Glu Leu Leu Gly Lys Arg Ala Leu Gly Asn
      1470                1475                1480

AAT TCA AGG GAT GTA AAA TCA ACA AGA AAA CTT ACT CGG GCA TGG CAT
Asn Ser Arg Asp Val Lys Ser Thr Arg Lys Leu Thr Arg Ala Trp His
      1485                1490                1495
```

FIG.6N.

```
GAT CAT TGT GAT TAT AAA GGT AAC TCC TCT AAT TAC AGA GAC TGT AAA
Asp His Cys Asp Tyr Lys Gly Asn Ser Ser Asn Tyr Arg Asp Cys Lys
        1215                1220                1225

GTG CGG TTA ATT AAA GGG AAA AAT TAT TAT TTC GCA GCA CGC AAT AAT
Val Arg Leu Ile Lys Gly Lys Asn Tyr Tyr Phe Ala Ala Arg Asn Asn
        1230                1235                1240

ATG GCA TTA GGG AAA TAC GTT GAT TTA GGT TTA GGT ATT CGG TAT GAC
Met Ala Leu Gly Lys Tyr Val Asp Leu Gly Leu Gly Ile Arg Tyr Asp
        1245                1250                1255

GTA TCT CGC ACA AAA GCT AAT GAA TCA ACT ATT AGT GTT GGT AAA TTT
Val Ser Arg Thr Lys Ala Asn Glu Ser Thr Ile Ser Val Gly Lys Phe
1260                1265                1270                1275

AAA AAT TTC TCT TGG AAT ACT GGT ATT GTC ATA AAA CCA ACG GAA TGG
Lys Asn Phe Ser Trp Asn Thr Gly Ile Val Ile Lys Pro Thr Glu Trp
        1280                1285                1290

CTT GAT CTT TCT TAT CGC CTT TCT ACT GGA TTT AGA AAT CCT AGT TTT
Leu Asp Leu Ser Tyr Arg Leu Ser Thr Gly Phe Arg Asn Pro Ser Phe
        1295                1300                1305
```

FIG.6L.

```
CGG GAT GAT AGC AGT GGC TCT TTT TAT CCA AAG CAA GAT TAT GGT GCA
Arg Asp Asp Ser Ser Gly Ser Phe Tyr Pro Lys Gln Asp Tyr Gly Ala
1020                1025                1030                1035

TAT CAA CGT ATT GAG GAT GGC CGA GGC GTT AAC TAT GCA AGT GGG CTT
Tyr Gln Arg Ile Glu Asp Gly Arg Gly Val Asn Tyr Ala Ser Gly Leu
                    1040                1045                1050

TAT TTC GAT GAA CAC CAT AGA AAA CAG CGT GTA GGT ATT GAA TAT ATT
Tyr Phe Asp Glu His His Arg Lys Gln Arg Val Gly Ile Glu Tyr Ile
                    1055                1060                1065

TAC GAA AAT AAG AAC AAA GCG GGC ATC ATT GAC AAA GCA GTG TTA AGT
Tyr Glu Asn Lys Asn Lys Ala Gly Ile Ile Asp Lys Ala Val Leu Ser
                    1070                1075                1080

GCT AAT CAA CAA AAC ATC ATA CTT GAC AGT TAT ATG CAA CAT ACG CAT
Ala Asn Gln Gln Asn Ile Ile Leu Asp Ser Tyr Met Gln His Thr His
                    1085                1090                1095

TGC AGT CTT TAT CCT AAT CCA AGT AAG AAT TGC CGC CCA ACA CGT GAT
Cys Ser Leu Tyr Pro Asn Pro Ser Lys Asn Cys Arg Pro Thr Arg Asp
1100                1105                1110                1115
```

FIG.6J.

```
TAT GGT AAT GGA GCA CTA GCT GGT TCT GTA ACA TTT CAA AGC AAA TCA
Tyr Gly Asn Gly Ala Leu Ala Gly Ser Val Thr Phe Gln Ser Lys Ser
          830                 835                 840

GCA GCC GAT ATC TTA GAA GGA GAC AAA TCA TGG GGA ATT CAA ACT AAA
Ala Ala Asp Ile Leu Glu Gly Asp Lys Ser Trp Gly Ile Gln Thr Lys
          845                 850                 855

AAT GCT TAT TCA AGC AAA AAT AAA GGC TTT ACC CAT TCT TTA GCT GTA
Asn Ala Tyr Ser Ser Lys Asn Lys Gly Phe Thr His Ser Leu Ala Val
          860                 865                 870                 875

GCT GGA AAA CAA GGG GGA TTT GAC GGG GTC GCC ATT TAT ACT CAA CGA
Ala Gly Lys Gln Gly Gly Phe Asp Gly Val Ala Ile Tyr Thr Gln Arg
                    880                 885                 890

AAT TCA ATT GAA ACC CAA GTC CAT AAA GAT GCA TTA AAA GGC GTA CAA
Asn Ser Ile Glu Thr Gln Val His Lys Asp Ala Leu Lys Gly Val Gln
          895                 900                 905

AGT TAT CAT CGA TTA ATC GCC AAA CCA GAG GAT CAA TCT GCA TAC TTT
Ser Tyr His Arg Leu Ile Ala Lys Pro Glu Asp Gln Ser Ala Tyr Phe
          910                 915                 920
```

FIG.6H.

```
AAA AAA CAA GTA GAA ACA ACC AAC AAG TAAAAACAAC CAAGTAATGG
Lys Lys Gln Val Glu Thr Thr Asn Lys
                650

AATACTAAAA ATG ACT AAA AAA CCC TAT TTT CGC CTA AGT ATT ATT TCT
           Met Thr Lys Lys Pro Tyr Phe Arg Leu Ser Ile Ile Ser
           655                 660                 665

TGT CTT TTA ATT TCA TGC TAT GTA AAA GCA GAA ACT CAA AGT ATA AAA
Cys Leu Leu Ile Ser Cys Tyr Val Lys Ala Glu Thr Gln Ser Ile Lys
        670                 675                 680

GAT ACA AAA GAA GCT ATA TCA TCT GAA GTG GAC ACT CAA AGT ACA GAA
Asp Thr Lys Glu Ala Ile Ser Ser Glu Val Asp Thr Gln Ser Thr Glu
    685                 690                 695

GAT TCA GAA TTA GAA ACT ATC TCA GTC ACT GCA GAA AAA ATA AGA GAT
Asp Ser Glu Leu Glu Thr Ile Ser Val Thr Ala Glu Lys Ile Arg Asp
700                 705                 710                 715

CGT AAA GAT AAT GAA GTA ACT GGA CTT GGC AAA ATT ATC AAA ACT AGT
Arg Lys Asp Asn Glu Val Thr Gly Leu Gly Lys Ile Ile Lys Thr Ser
                720                 725                 730
```

FIG.6F.

```
GAC AAC AAA AAT GAA ACA GAC AAA GAA AAA GGC AAA GAA AAA CCA ACG
Asp Asn Lys Asn Glu Thr Asp Lys Glu Lys Gly Lys Glu Lys Pro Thr
 455                 460                 465

ACG ACA ACA TCT ATC AAC ACT TAT TAT CAA TTC TTA TTA GGT CTC CGT
Thr Thr Thr Ser Ile Asn Thr Tyr Tyr Gln Phe Leu Leu Gly Leu Arg
 470                 475                 480                 485

ACT CCC AAG GAC GAA ATA CCT AAA GAA GGA AGT GCA AAA TAT CAT GGT
Thr Pro Lys Asp Glu Ile Pro Lys Glu Gly Ser Ala Lys Tyr His Gly
                 490                 495                 500

AAT TGG TTT GGT TAT ATT AGT GAT GGC GAG ACA TCT TAC TCC GCC AGT
Asn Trp Phe Gly Tyr Ile Ser Asp Gly Glu Thr Ser Tyr Ser Ala Ser
                 505                 510                 515

GGT GAT AAG GAA CGC AGT AAA AAT GCT GTC GCC GAG TTT GAT GTA AGT
Gly Asp Lys Glu Arg Ser Lys Asn Ala Val Ala Glu Phe Asp Val Ser
                 520                 525                 530

TTT GCC AAT AAA ACA TTA ACA GGC GAA TTA AAA CGA CAC GAT AAT GGA
Phe Ala Asn Lys Thr Leu Thr Gly Glu Leu Lys Arg His Asp Asn Gly
                 535                 540                 545
```

FIG.6D.

```
CTC TAC ACT CTA GAA GCT AAA GTG TAT AGC AAC CGA TTC AGA GGT AAA
Leu Tyr Thr Leu Glu Ala Lys Val Tyr Ser Asn Arg Phe Arg Gly Lys
                265                 270                 275

GTA AAG CCA ACC AAA ACA AAG TCT GAA GAT CAT CCC TTT ACC AGC GAG
Val Lys Pro Thr Lys Thr Lys Ser Glu Asp His Pro Phe Thr Ser Glu
            280                 285                 290

GGA ACA TTA GAA GGT GGT TTT TAT GGG CCT AAT GCT GAA GAA CTA GGG
Gly Thr Leu Glu Gly Gly Phe Tyr Gly Pro Asn Ala Glu Glu Leu Gly
        295                 300                 305

GGA AAG TTT TTA GCT AAC GAC GAA AAA GTT TTT GGG GTA TTT AGT GCC
Gly Lys Phe Leu Ala Asn Asp Glu Lys Val Phe Gly Val Phe Ser Ala
    310                 315                 320                 325

AAA GAA GAC CCA CAA AAC CCA GAA AAC CAA AAA TTA TCC ACA GAA ACC
Lys Glu Asp Pro Gln Asn Pro Glu Asn Gln Lys Leu Ser Thr Glu Thr
                330                 335                 340

TTA ATT GAT GGC AAG CTA ATT ACT TTT AAA AGA ACT GAT GCA ACA ACC
Leu Ile Asp Gly Lys Leu Ile Thr Phe Lys Arg Thr Asp Ala Thr Thr
            345                 350                 355
```

FIG.6B.

```
CAA AAT TTT ATT GGT GCT AGA GAA CCT AGT TTC TTA AAT GAA GAT GGC
Gln Asn Phe Ile Gly Ala Arg Glu Pro Ser Phe Leu Asn Glu Asp Gly
 70                  75                  80                  85

TAT ATG ATA TTT TCC TCA CTT TCT ACG ATT GAA GAG GAT GTT GAA AAA
Tyr Met Ile Phe Ser Ser Leu Ser Thr Ile Glu Glu Asp Val Glu Lys
                 90                  95                 100

GTT AAA AAT AAC AAT AAA AAC GGG GGG AGG CTT ATT GGC TCA ATT GAG
Val Lys Asn Asn Asn Lys Asn Gly Gly Arg Leu Ile Gly Ser Ile Glu
                105                 110                 115

GAA CCT AAT GGA ACA TCA CAA AAT TCT AAT TCA CAA GAA TAC GTT TAT
Glu Pro Asn Gly Thr Ser Gln Asn Ser Asn Ser Gln Glu Tyr Val Tyr
                120                 125                 130

TCT GGT TTG TAT TAT ATC GAT AGT TGG CGT GAT TAT AAG AAG GAA GAG
Ser Gly Leu Tyr Tyr Ile Asp Ser Trp Arg Asp Tyr Lys Lys Glu Glu
                135                 140                 145

CAA AAA GCT TAT ACT GGC TAT TAT GGT TAT GCA TTT TAT TAT GGT AAT
Gln Lys Ala Tyr Thr Gly Tyr Tyr Gly Tyr Ala Phe Tyr Tyr Gly Asn
150                 155                 160                 165
```

FIG.5Q.

```
AAT ATT ATG CTT CGA TTA GGG ATA TAT AAT TTA TTC AAC TAT CGC TAT
Asn Ile Met Leu Arg Leu Gly Ile Tyr Asn Leu Phe Asn Tyr Arg Tyr
                1520                1525                1530

GTT ACT TGG GAA GCG GTG CGT CAA ACA GCA CAA GGT GCG GTC AAT CAA
Val Thr Trp Glu Ala Val Arg Gln Thr Ala Gln Gly Ala Val Asn Gln
            1535                1540                1545

CAT CAA AAT GTT GGT AGC TAT ACT CGC TAC GCA GCA TCA GGA CGA AAC
His Gln Asn Val Gly Ser Tyr Thr Arg Tyr Ala Ala Ser Gly Arg Asn
        1550                1555                1560

TAT ACC TTA ACA TTA GAA ATG AAA TTC TAAATTAAAA TGCGCCAGAT
Tyr Thr Leu Thr Leu Glu Met Lys Phe
    1565                1570

GGACTAGATA TGCTATATCT ATACCTTACT GGCGCATCTT TTTCTGTTCT ATAATCTGCT

TAAGTGAAAA ACCAAACTTG GATTTTTTAC AAGATCTTTT CACACATTTA TTGTAAAATC

TCCGACAATT TTGACCG
```

FIG.50.

```
GAC GAG GTT TAT GTA GGT AAA TTT AAG CCT GAA ACA TCT CGT AAC CAA
Asp Glu Val Tyr Val Gly Lys Phe Lys Pro Glu Thr Ser Arg Asn Gln
    1325                1330                1335

GAG TTT GGT CTC GCT CTA AAA GGG GAT TTT GGT AAT ATT GAG ATC AGT
Glu Phe Gly Leu Ala Leu Lys Gly Asp Phe Gly Asn Ile Glu Ile Ser
1340                1345                1350                1355

CAT TTT AGT AAT GCT TAT CGA AAT CTT ATC GCC TTT GCT GAA GAA CTT
His Phe Ser Asn Ala Tyr Arg Asn Leu Ile Ala Phe Ala Glu Glu Leu
                1360                1365                1370

AGT AAA AAT GGA ACT GGA AAG GGC AAT TAT GGA TAT CAT AAT GCA CAA
Ser Lys Asn Gly Thr Gly Lys Gly Asn Tyr Gly Tyr His Asn Ala Gln
            1375                1380                1385

AAT GCA AAA TTA GTT GGC GTA AAT ATA ACT GCA CAA TTA GAT TTT AAT
Asn Ala Lys Leu Val Gly Val Asn Ile Thr Ala Gln Leu Asp Phe Asn
        1390                1395                1400

GGT TTA TGG AAA CGT ATT CCC TAC GGT TGG TAT GCA ACA TTT GCT TAT
Gly Leu Trp Lys Arg Ile Pro Tyr Gly Trp Tyr Ala Thr Phe Ala Tyr
    1405                1410                1415
```

FIG.5M.

```
GAA AAA CAT AAT ATG TTG CAA TTG AAT TTA GAG AAA AAA ATT CAA CAA
Glu Lys His Asn Met Leu Gln Leu Asn Leu Glu Lys Lys Ile Gln Gln
                1135                1140                1145

AAT TGG CTT ACT CAT CAA ATT GTC TTC AAT CTT GGT TTT GAT GAC TTT
Asn Trp Leu Thr His Gln Ile Val Phe Asn Leu Gly Phe Asp Asp Phe
                1150                1155                1160

ACT TCA GCG CTT CAG CAT AAA GAT TAT TTA ACT CGA CGT GTT ATC GCT
Thr Ser Ala Leu Gln His Lys Asp Tyr Leu Thr Arg Arg Val Ile Ala
                1165                1170                1175

ACG GCA GAT AGT ATT CCA AGG AAA CCT GGT GAA ACT GGT AAA CCA AGA
Thr Ala Asp Ser Ile Pro Arg Lys Pro Gly Glu Thr Gly Lys Pro Arg
1180                1185                1190                1195

AAT GGT TTG CAA TCA CAA CCT TAC TTA TAC CCA AAA CCA GAG CCA TAT
Asn Gly Leu Gln Ser Gln Pro Tyr Leu Tyr Pro Lys Pro Glu Pro Tyr
                1200                1205                1210

TTT GCA GGA CAA GAT CAT TGT AAT TAT CAA GGT AGC TCC TCT AAT TAC
Phe Ala Gly Gln Asp His Cys Asn Tyr Gln Gly Ser Ser Ser Asn Tyr
                1215                1220                1225
```

FIG.5K.

```
GAC AAG TGT GCA GCC AAG CCA CCT GCG ACT TTA TCC ACC CAA AGC GAA
Asp Lys Cys Ala Ala Lys Pro Pro Ala Thr Leu Ser Thr Gln Ser Glu
940                 945                 950                 955

ACC GTA AGC GTT TCA GAT TAT ACG GGG GCT AAC CGT ATC AAA CCT AAT
Thr Val Ser Val Ser Asp Tyr Thr Gly Ala Asn Arg Ile Lys Pro Asn
                960                 965                 970

CCA ATG AAA TAT GAA AGC CAG TCT TGG TTT TTA AGA GGA GGG TAT CAT
Pro Met Lys Tyr Glu Ser Gln Ser Trp Phe Leu Arg Gly Gly Tyr His
            975                 980                 985

TTT TCT GAA CAA CAT TAT ATT GGT GGT ATT TTT GAA TTC ACA CAA CAA
Phe Ser Glu Gln His Tyr Ile Gly Gly Ile Phe Glu Phe Thr Gln Gln
        990                 995                1000

AAA TTT GAT ATC CGT GAT ATG ACA TTT CCC GCT TAT TTA AGC CCA ACA
Lys Phe Asp Ile Arg Asp Met Thr Phe Pro Ala Tyr Leu Ser Pro Thr
    1005                1010                1015

GAA AGA CGG GAT GAT AGT AGT CGT TCT TTT TAT CCA ATG CAA GAT CAT
Glu Arg Arg Asp Asp Ser Ser Arg Ser Phe Tyr Pro Met Gln Asp His
1020                1025                1030                1035
```

FIG.5H.

```
GAA ACA ACC AAA TAATGGAATA CTAAAA ATG ACT AAA AAA CCC TAT TTT
Glu Thr Thr Lys                Met Thr Lys Lys Pro Tyr Phe
660                                    665

CGC CTA AGT ATT ATT TCT TGT CTT TTA ATT TCA TGC TAT GTA AAA GCA
Arg Leu Ser Ile Ile Ser Cys Leu Leu Ile Ser Cys Tyr Val Lys Ala
670                    675                 680

GAA ACT CAA AGT ATA AAA GAT ACA AAA GAA GCT ATA TCA TCT GAA GTG
Glu Thr Gln Ser Ile Lys Asp Thr Lys Glu Ala Ile Ser Ser Glu Val
685                    690                 695

GAC ACT CAA AGT ACA GAA GAT TCA GAA TTA GAA ACT ATC TCA GTC ACT
Asp Thr Gln Ser Thr Glu Asp Ser Glu Leu Glu Thr Ile Ser Val Thr
700                    705                 710                 715

GCA GAA AAA ATA AGA GAT CGT AAA GAT AAT GAA GTA ACT GGA CTT GGC
Ala Glu Lys Ile Arg Asp Arg Lys Asp Asn Glu Val Thr Gly Leu Gly
                720                 725                 730

AAA ATT ATC AAA ACT AGT GAA AGT ATC AGC CGA GAA CAA GTA TTA AAT
Lys Ile Ile Lys Thr Ser Glu Ser Ile Ser Arg Glu Gln Val Leu Asn
                735                 740                 745
```

FIG.5F.

```
AAA GAA AAA GAA AAA GAC AAA GAC AAA GAA AAA CAA ACG GCG GCA ACG
Lys Glu Lys Glu Lys Asp Lys Asp Lys Glu Lys Gln Thr Ala Ala Thr
465                 470                 475                 480

ACC AAC ACT TAT TAT CAA TTC TTA TTA GGT CAC CGT ACT CCC AAG GAC
Thr Asn Thr Tyr Tyr Gln Phe Leu Leu Gly His Arg Thr Pro Lys Asp
                485                 490                 495

GAC ATA CCT AAA ACA GGA AGT GCA AAA TAT CAT GGT AGT TGG TTT GGT
Asp Ile Pro Lys Thr Gly Ser Ala Lys Tyr His Gly Ser Trp Phe Gly
            500                 505                 510

TAT ATT ACT GAC GGT AAG ACA TCT TAC TCC CCC AGT GGT GAT AAG AAA
Tyr Ile Thr Asp Gly Lys Thr Ser Tyr Ser Pro Ser Gly Asp Lys Lys
        515                 520                 525

CGC GAT AAA AAT GCT GTC GCC GAG TTT AAT GTT GAT TTT GCC GAG AAA
Arg Asp Lys Asn Ala Val Ala Glu Phe Asn Val Asp Phe Ala Glu Lys
    530                 535                 540

AAG CTA ACA GGC GAA TTA AAA CGA CAC GAT ACT GGA AAT CCC GTA TTT
Lys Leu Thr Gly Glu Leu Lys Arg His Asp Thr Gly Asn Pro Val Phe
545                 550                 555                 560
```

FIG.5D.

```
AGG GGT ACA GTA AAG CCA ACC GAA AAA GAT TCT GAA GAA CAT CCC TTT
Arg Gly Thr Val Lys Pro Thr Glu Lys Asp Ser Glu Glu His Pro Phe
        275                 280                 285

ACC AGC GAG GGA ACA TTA GAA GGT GGT TTT TAT GGG CCT AAT GCT GAA
Thr Ser Glu Gly Thr Leu Glu Gly Gly Phe Tyr Gly Pro Asn Ala Glu
    290                 295                 300

GAA CTA GGG GGG AAA TTT TTA GCT ACG GAT AAC CGA GTT TTT GGG GTA
Glu Leu Gly Gly Lys Phe Leu Ala Thr Asp Asn Arg Val Phe Gly Val
305                 310                 315                 320

TTT AGT GCC AAA GAA ACG GAA GAA ACA AAA AAG GAA GCG TTA TCC AAG
Phe Ser Ala Lys Glu Thr Glu Glu Thr Lys Lys Glu Ala Leu Ser Lys
                325                 330                 335

GAA ACC TTA ATT GAT GGC AAG CTA ATT ACT TTC TCT ACT AAA AAA ACC
Glu Thr Leu Ile Asp Gly Lys Leu Ile Thr Phe Ser Thr Lys Lys Thr
            340                 345                 350

GAT GCA AAA ACC AAT GCA ACA ACC AGT ACC GCA GCT AAT ACA ACA ACC
Asp Ala Lys Thr Asn Ala Thr Thr Ser Thr Ala Ala Asn Thr Thr Thr
        355                 360                 365
```

FIG.5B.

```
AAT GAA GAT GAC TAT ATA TCA TAT TTT TCC TCA CTT TCT ACG ATT GAA
Asn Glu Asp Asp Tyr Ile Ser Tyr Phe Ser Ser Leu Ser Thr Ile Glu
                85                  90                  95

AAG GAT GTT AAA GAT AAC AAT AAA AAC GGG GCG GAC CTT ATT GGC TCA
Lys Asp Val Lys Asp Asn Asn Lys Asn Gly Ala Asp Leu Ile Gly Ser
            100                 105                 110

ATA GAC GAG CCT AGT ACA ACA AAT CCA CCC GAA AAG CAT CAT GGA CAA
Ile Asp Glu Pro Ser Thr Thr Asn Pro Pro Glu Lys His His Gly Gln
        115                 120                 125

AAA TAT GTA TAT TCA GGG CTT TAT TAT ACT CCA TCG TGG AGT TTA AAC
Lys Tyr Val Tyr Ser Gly Leu Tyr Tyr Thr Pro Ser Trp Ser Leu Asn
    130                 135                 140

GAT TCT AAA AAC AAG TTT TAT TTA GGT TAC TAT GGA TAT GCG TTT TAT
Asp Ser Lys Asn Lys Phe Tyr Leu Gly Tyr Tyr Gly Tyr Ala Phe Tyr
145                 150                 155                 160

TAT GGT AAT AAA ACT GCA ACA AAC TTG CCA GTA AAC GGT GTA GCT AAA
Tyr Gly Asn Lys Thr Ala Thr Asn Leu Pro Val Asn Gly Val Ala Lys
                165                 170                 175
```

FIG.4Q.

```
CAT ATC TTA GAT GTA TCG GGT TAT TAC ATG GCG AAT AAA AAT ATT ATG
His Ile Leu Asp Val Ser Gly Tyr Tyr Met Ala Asn Lys Asn Ile Met
        1505                1510                1515

CTT CGA TTA GGG ATA TAT AAT TTA TTC AAC TAT CGC TAT GTT ACT TGG
Leu Arg Leu Gly Ile Tyr Asn Leu Phe Asn Tyr Arg Tyr Val Thr Trp
    1520                1525                1530

GAA GCG GTG CGT CAA ACA GCA CAA GGT GCG GTC AAT CAA CAT CAA AAT
Glu Ala Val Arg Gln Thr Ala Gln Gly Ala Val Asn Gln His Gln Asn
1535                1540                1545                1550

GTT GGT AGC TAT ACT CGC TAC GCA GCA TCA GGA CGA AAC TAT ACC TTA
Val Gly Ser Tyr Thr Arg Tyr Ala Ala Ser Gly Arg Asn Tyr Thr Leu
                1555                1560                1565

ACA TTA GAA ATG AAA TTC TAAATTAAAA TGCGCCAGAT GGACTAGATA
Thr Leu Glu Met Lys Phe
            1570

TGCTATATCT ATACCTTACT GGCGCATCTT TTTCTGTTCT ATAATCTGCT TAAGTGAAAA

ACCAAACTTG GATTTTTTAC AAGATCTTTT CACACATTTA TTG
```

FIG.40.

```
TTT TCT GAA ATG TAT GGT TGG CGG TAT GGT GGC AAG AAT GAC GAG GTT
Phe Ser Glu Met Tyr Gly Trp Arg Tyr Gly Gly Lys Asn Asp Glu Val
                1315                1320                1325

TAT GTA GGT AAA TTT AAG CCT GAA ACA TCT CGT AAC CAA GAG TTT GGT
Tyr Val Gly Lys Phe Lys Pro Glu Thr Ser Arg Asn Gln Glu Phe Gly
                1330                1335                1340

CTC GCT CTA AAA GGG GAT TTT GGT AAT ATT GAG ATC AGT CAT TTT AGT
Leu Ala Leu Lys Gly Asp Phe Gly Asn Ile Glu Ile Ser His Phe Ser
                1345                1350                1355

AAT GCT TAT CGA AAT CTT ATC GCC TTT GCT GAA GAA CTT AGT AAA AAT
Asn Ala Tyr Arg Asn Leu Ile Ala Phe Ala Glu Glu Leu Ser Lys Asn
                1360                1365                1370

GGA ACT GGA AAG GGC AAT TAT GGA TAT CAT AAT GCA CAA AAT GCA AAA
Gly Thr Gly Lys Gly Asn Tyr Gly Tyr His Asn Ala Gln Asn Ala Lys
1375                1380                1385                1390

TTA GTT GGC GTA AAT ATA ACT GCA CAA TTA GAT TTT AAT GGT TTA TGG
Leu Val Gly Val Asn Ile Thr Ala Gln Leu Asp Phe Asn Gly Leu Trp
                1395                1400                1405
```

FIG.4M.

```
CCT TAT TCA TAC TAT CGT TCT GAT AGA AAT GTT TAT AAA GAA AAA CAT
Pro Tyr Ser Tyr Tyr Arg Ser Asp Arg Asn Val Tyr Lys Glu Lys His
        1120                1125                1130

AAT ATG TTG CAA TTG AAT TTA GAG AAA AAA ATT CAA CAA AAT TGG CTT
Asn Met Leu Gln Leu Asn Leu Glu Lys Lys Ile Gln Gln Asn Trp Leu
1135                1140                1145                1150

ACT CAT CAA ATT GTC TTC AAT CTT GGT TTT GAT GAC TTT ACT TCA GCG
Thr His Gln Ile Val Phe Asn Leu Gly Phe Asp Asp Phe Thr Ser Ala
                1155                1160                1165

CTT CAG CAT AAA GAT TAT TTA ACT CGA CGT GTT ATC GCT ACG GCA GAT
Leu Gln His Lys Asp Tyr Leu Thr Arg Arg Val Ile Ala Thr Ala Asp
        1170                1175                1180

AGT ATT CCA AGG AAA CCT GGT GAA ACT GGT AAA CCA AGA AAT GGT TTG
Ser Ile Pro Arg Lys Pro Gly Glu Thr Gly Lys Pro Arg Asn Gly Leu
        1185                1190                1195

CAA TCA CAA CCT TAC TTA TAC CCA AAA CCA GAG CCA TAT TTT GCA GGA
Gln Ser Gln Pro Tyr Leu Tyr Pro Lys Pro Glu Pro Tyr Phe Ala Gly
        1200                1205                1210
```

FIG.4K.

```
GGA TAC TTT GTG ATA CAA GGT GAG TGT CCA AAT GGT GAT GAC AAG TGT
Gly Tyr Phe Val Ile Gln Gly Glu Cys Pro Asn Gly Asp Asp Lys Cys
          930                 935                 940

GCA GCC AAG CCA CCT GCG ACT TTA TCC ACC CAA AGC GAA ACC GTA AGC
Ala Ala Lys Pro Pro Ala Thr Leu Ser Thr Gln Ser Glu Thr Val Ser
          945                 950                 955

GTT TCA GAT TAT ACG GGG GCT AAC CGT ATC AAA CCT AAT CCA ATG AAA
Val Ser Asp Tyr Thr Gly Ala Asn Arg Ile Lys Pro Asn Pro Met Lys
          960                 965                 970

TAT GAA AGC CAG TCT TGG TTT TTA AGA GGA GGG TAT CAT TTT TCT GAA
Tyr Glu Ser Gln Ser Trp Phe Leu Arg Gly Gly Tyr His Phe Ser Glu
          975                 980                 985                 990

CAA CAT TAT ATT GGT GGT ATT TTT GAA TTC ACA CAA CAA AAA TTT GAT
Gln His Tyr Ile Gly Gly Ile Phe Glu Phe Thr Gln Gln Lys Phe Asp
          995                 1000                1005

ATC CGT GAT ATG ACA TTT CCC GCT TAT TTA AGC CCA ACA GAA AGA CGG
Ile Arg Asp Met Thr Phe Pro Ala Tyr Leu Ser Pro Thr Glu Arg Arg
          1010                1015                1020
```

FIG.41.

```
AAA ACT AGT GAA AGT ATC AGC CGA GAA CAA GTA TTA AAT ATT CGT GAT
Lys Thr Ser Glu Ser Ile Ser Arg Glu Gln Val Leu Asn Ile Arg Asp
735                 740                 745                 750

CTA ACA CGC TAT GAT CCA GGG ATT TCA GTT GTA GAA CAA GGT CGC GGT
Leu Thr Arg Tyr Asp Pro Gly Ile Ser Val Val Glu Gln Gly Arg Gly
                755                 760                 765

GCA AGT TCT GGA TAT TCT ATT CGT GGT ATG GAC AGA AAT AGA GTT GCT
Ala Ser Ser Gly Tyr Ser Ile Arg Gly Met Asp Arg Asn Arg Val Ala
            770                 775                 780

TTA TTA GTA GAT GGT TTA CCT CAA ACG CAA TCT TAT GTA GTG CAA AGC
Leu Leu Val Asp Gly Leu Pro Gln Thr Gln Ser Tyr Val Val Gln Ser
        785                 790                 795

CCT TTA GTT GCT CGT TCA GGA TAT TCT GGC ACT GGT GCA ATT AAT GAA
Pro Leu Val Ala Arg Ser Gly Tyr Ser Gly Thr Gly Ala Ile Asn Glu
    800                 805                 810

ATT GAA TAT GAA AAT GTA AAG GCC GTC GAA ATA AGC AAG GGG GGG AGT
Ile Glu Tyr Glu Asn Val Lys Ala Val Glu Ile Ser Lys Gly Gly Ser
815                 820                 825                 830
```

FIG.4G.

```
GGC GAA TTA AAA CGA CAC GAT ACT GGA AAT CCC GTA TTT AGT ATT GAG
Gly Glu Leu Lys Arg His Asp Thr Gly Asn Pro Val Phe Ser Ile Glu
        550                 555                 560

GCA AAC TTT AAT AAT AGT AGT AAT GCC TTC ACT GGT ACA GCA ACC GCA
Ala Asn Phe Asn Asn Ser Ser Asn Ala Phe Thr Gly Thr Ala Thr Ala
    565                 570                 575

ACA AAT TTT GTA ATA GAT GGT AAA AAT AGT CAA AAT AAA AAT ACC CCA
Thr Asn Phe Val Ile Asp Gly Lys Asn Ser Gln Asn Lys Asn Thr Pro
580                 585                 590                 595

ATT AAT ATT ACA ACT AAA GTA AAC GGG GCA TTT TAT GGA CCT AAG GCT
Ile Asn Ile Thr Thr Lys Val Asn Gly Ala Phe Tyr Gly Pro Lys Ala
                600                 605                 610

TCT GAA TTA GGC GGT TAT TTC ACT TAT AAC GGA AAT TCT ACA GCT ACA
Ser Glu Leu Gly Gly Tyr Phe Thr Tyr Asn Gly Asn Ser Thr Ala Thr
            615                 620                 625

AAT TCT GAA AGT TCC TCA ACC GTA TCT TCA TCA TCC AAT TCA AAA AAT
Asn Ser Glu Ser Ser Ser Thr Val Ser Ser Ser Ser Asn Ser Lys Asn
        630                 635                 640
```

FIG.4E.

```
ACC AAT GCA ACA ACC AGT ACC GCA GCT AAT ACA ACA ACC GAT ACA ACC
Thr Asn Ala Thr Thr Ser Thr Ala Ala Asn Thr Thr Thr Asp Thr Thr
          360                 365                 370

GCC AAT ACA ATA ACC GAT GAA AAA AAC TTT AAG ACG GAA GAT ATA TCA
Ala Asn Thr Ile Thr Asp Glu Lys Asn Phe Lys Thr Glu Asp Ile Ser
          375                 380                 385

AGT TTT GGT GAA GCT GAT TAT CTG TTA ATT GAC AAA TAT CCT ATT CCA
Ser Phe Gly Glu Ala Asp Tyr Leu Leu Ile Asp Lys Tyr Pro Ile Pro
          390                 395                 400

CTT TTA CCT GAT AAA AAT ACT AAT GAT TTC ATA AGT AGT AAG CAT CAT
Leu Leu Pro Asp Lys Asn Thr Asn Asp Phe Ile Ser Ser Lys His His
          405                 410                 415

ACT GTA GGA AAT AAA CGC TAT AAA GTG GAA GCA TGT TGC AGT AAT CTA
Thr Val Gly Asn Lys Arg Tyr Lys Val Glu Ala Cys Cys Ser Asn Leu
          420                 425                 430                 435

AGC TAT GTG AAA TTT GGT ATG TAT TAT GAA GAC CCA CTT AAA GAA AAA
Ser Tyr Val Lys Phe Gly Met Tyr Tyr Tyr Glu Asp Pro Leu Lys Glu Lys
          440                 445                 450
```

FIG.4C.

```
AAA ACT GCA ACA AAC TTG CCA GTA AAC GGT GTA GCT AAA TAC AAA GGA
Lys Thr Ala Thr Asn Leu Pro Val Asn Gly Val Ala Lys Tyr Lys Gly
      165                 170                 175

ACT TGG GAT TTC ATC ACT GCA ACT AAA AAT GGC AAA CGT TAT CCT TTG
Thr Trp Asp Phe Ile Thr Ala Thr Lys Asn Gly Lys Arg Tyr Pro Leu
  180                 185                 190                 195

TTA AGT AAT GGC AGT CAC GCT TAT TAT CGA CGT AGT GCA ATT CCA GAA
Leu Ser Asn Gly Ser His Ala Tyr Tyr Arg Arg Ser Ala Ile Pro Glu
                  200                 205                 210

GAT ATT GAT TTA GAA AAT GAT TCA AAG AAT GGT GAT ATA GGC TTA ATA
Asp Ile Asp Leu Glu Asn Asp Ser Lys Asn Gly Asp Ile Gly Leu Ile
              215                 220                 225

AGT GAA TTT AGT GCA GAT TTT GGG ACT AAA AAA CTG ACA GGA CAA CTG
Ser Glu Phe Ser Ala Asp Phe Gly Thr Lys Lys Leu Thr Gly Gln Leu
          230                 235                 240

TCT TAC ACC AAA AGA AAA ACT AAT AAT CAA CCA TAT GAA AAG AAA AAA
Ser Tyr Thr Lys Arg Lys Thr Asn Asn Gln Pro Tyr Glu Lys Lys Lys
      245                 250                 255
```

FIG.4A.

```
GCCCAAGCTA CATTGGTTAA TGATAAGCCT ATAAATGATA AGAAAGAAAT TTGTTTTACG

CCATTTTTCA TATTTTATCC ATGAACTTAA AAAACTCTAA CTTGACATTA TTACAAAAAA

AGATCAATAA TGCGAATTAT TATCAATTTT GTATGAGTAT ATAATTCT ATG AAA TCT
                                                  Met Lys Ser
                                                    1

GTA CCT CTT ATC TCT GGT GGA CTT TCC TTT TTA CTA AGT GCT TGT AGC
Val Pro Leu Ile Ser Gly Gly Leu Ser Phe Leu Leu Ser Ala Cys Ser
     5                  10                  15

GGA GGG GGG TCT TTT GAT GTA GAT AAC GTC TCT AAT ACC CCC TCT TCT
Gly Gly Gly Ser Phe Asp Val Asp Asn Val Ser Asn Thr Pro Ser Ser
 20                  25                  30                  35

AAA CCA CGT TAT CAA GAC GAT ACC TCG AAT CAA AGA AAA AAA TCT AAT
Lys Pro Arg Tyr Gln Asp Asp Thr Ser Asn Gln Arg Lys Lys Ser Asn
                 40                  45                  50

TTG AAA AAG TTG TTC ATT CCT TCT TTA GGA GGA GGG ATG AAA TTG GTG
Leu Lys Lys Leu Phe Ile Pro Ser Leu Gly Gly Gly Met Lys Leu Val
            55                  60                  65
```

FIG.3P.

```
CGT TAT ATC ATT GGT TTA GGC TAT GAT CAT CCA AGT AAT ACT TGG GGA
Arg Tyr Ile Ile Gly Leu Gly Tyr Asp His Pro Ser Asn Thr Trp Gly
        1435                1440                1445

ATT AAG ACA ATG TTT ACT CAA TCA AAA GCA AAA TCT CAA AAT GAA TTG
Ile Lys Thr Met Phe Thr Gln Ser Lys Ala Lys Ser Gln Asn Glu Leu
        1450                1455                1460

CTA GGA AAA CGT GCA TTG GGT AAC AAT TCA AGG AAT GTA AAA TCA ACA
Leu Gly Lys Arg Ala Leu Gly Asn Asn Ser Arg Asn Val Lys Ser Thr
1465                1470                1475                1480

AGA AAA CTT ACT CGG GCA TGG CAT ATC TTA GAT GTA TCG GGT TAT TAC
Arg Lys Leu Thr Arg Ala Trp His Ile Leu Asp Val Ser Gly Tyr Tyr
                1485                1490                1495

ATG GTG AAT AGA AGT ATT TTG TTC CGA TTA GGA GTA TAT AAT TTA TTA
Met Val Asn Arg Ser Ile Leu Phe Arg Leu Gly Val Tyr Asn Leu Leu
        1500                1505                1510

AAC TAT CGC TAT GTC ACT TGG GAA GCG GTG CGT CAA ACA GCA CAA GGT
Asn Tyr Arg Tyr Val Thr Trp Glu Ala Val Arg Gln Thr Ala Gln Gly
        1515                1520                1525
```

FIG. 3N.

```
GGT TTA GGT ATG AGG TAT GAC GTA TCT CGT ACA AAA GCT AAT GAA TCA
Gly Leu Gly Met Arg Tyr Asp Val Ser Arg Thr Lys Ala Asn Glu Ser
                1245                1250                1255

ACT ATT AGT GTT GGT AAA TTT AAA AAT TTC TCT TGG AAT ACT GGT ATT
Thr Ile Ser Val Gly Lys Phe Lys Asn Phe Ser Trp Asn Thr Gly Ile
                1260                1265                1270

GTC ATA AAA CCA ACG GAA TGG CTT GAT CTT TCT TAT CGC CTT TCT ACT
Val Ile Lys Pro Thr Glu Trp Leu Asp Leu Ser Tyr Arg Leu Ser Thr
                1275                1280                1285

GGA TTT AGA AAT CCT AGT TTT GCT GAA ATG TAT GGT TGG CGG TAT GGT
Gly Phe Arg Asn Pro Ser Phe Ala Glu Met Tyr Gly Trp Arg Tyr Gly
                1290                1295                1300

GGC AAG GAT ACC GAT GTT TAT ATA GGT AAA TTT AAG CCT GAA ACA TCT
Gly Lys Asp Thr Asp Val Tyr Ile Gly Lys Phe Lys Pro Glu Thr Ser
1305                1310                1315                1320

CGT AAC CAA GAG TTT GGT CTC GCT CTA AAA GGG GAT TTT GGT AAT ATT
Arg Asn Gln Glu Phe Gly Leu Ala Leu Lys Gly Asp Phe Gly Asn Ile
                1325                1330                1335
```

FIG.3L.

```
CGT GTA GGT ATT GAA TAT ATT TAC GAA AAT AAG AAC AAA GCG GGC ATC
Arg Val Gly Ile Glu Tyr Ile Tyr Glu Asn Lys Asn Lys Ala Gly Ile
        1050                1055                1060

ATT GAC AAA GCG GTG TTA AGT GCT AAT CAA CAA ACA TCA TAC TTG ACA
Ile Asp Lys Ala Val Leu Ser Ala Asn Gln Gln Thr Ser Tyr Leu Thr
1065                1070                1075                1080

GTT ATA TGC GAC ATA CGC ATT GCA GTC TTT ATC CAT AAT CCA AGT AAG
Val Ile Cys Asp Ile Arg Ile Ala Val Phe Ile His Asn Pro Ser Lys
                1085                1090                1095

AAT TGC CGC CCA ACA CTT GAT AAA CCT TAT TCA TAC TAT CAT TCT GAT
Asn Cys Arg Pro Thr Leu Asp Lys Pro Tyr Ser Tyr Tyr His Ser Asp
        1100                1105                1110

AGA AAT GTT TAT AAA GAA AAA CAT AAC ATG TTG CAA TTG AAT TTA GAG
Arg Asn Val Tyr Lys Glu Lys His Asn Met Leu Gln Leu Asn Leu Glu
        1115                1120                1125

AAA AAA ATT CAA CAA AAT TGG CTT ACT CAT CAA ATT GCC TTC AAT CTT
Lys Lys Ile Gln Gln Asn Trp Leu Thr His Gln Ile Ala Phe Asn Leu
        1130                1135                1140
```

FIG.3J.

```
AAA GGC TTT ACC CAT TCT TTA GCT GTA GCA GGA AAA CAA GGT GGA TTT
Lys Gly Phe Thr His Ser Leu Ala Val Ala Gly Lys Gln Gly Gly Phe
                860                 865                 870

GAA GGG GTC GCC ATT TAC ACT CAC CGA AAT TCA ATT GAA ACC CAA GTC
Glu Gly Val Ala Ile Tyr Thr His Arg Asn Ser Ile Glu Thr Gln Val
            875                 880                 885

CAT AAA GAT GCA TTA AAA GGC GTG CAA AGT TAT GAT CGA TTC ATC GCC
His Lys Asp Ala Leu Lys Gly Val Gln Ser Tyr Asp Arg Phe Ile Ala
        890                 895                 900

ACA ACA GAG GAT CAA TCT GCA TAC TTT GTG ATG CAA GAT GAG TGT CTA
Thr Thr Glu Asp Gln Ser Ala Tyr Phe Val Met Gln Asp Glu Cys Leu
    905                 910                 915                 920

GAT GGT TAT GAC AAG TGT AAA ACT TCA CCC AAA CGA CCT GCG ACT TTA
Asp Gly Tyr Asp Lys Cys Lys Thr Ser Pro Lys Arg Pro Ala Thr Leu
                925                 930                 935

TCC ACC CAA AGA GAA ACC GTA AGC GTT TCA GAT TAT ACG GGG GCT AAC
Ser Thr Gln Arg Glu Thr Val Ser Val Ser Asp Tyr Thr Gly Ala Asn
            940                 945                 950
```

FIG.3H.

GTA AAA GCA GAA ACT CAA AGT ATA AAA GAT ACA AAA GAA GCT ATA TCA
Val Lys Ala Glu Thr Gln Ser Ile Lys Asp Thr Lys Glu Ala Ile Ser
665 670 675 680

TCT GAA GTG GAC ACT CAA AGT ACA GAA GAT TCA GAA TTA GAA ACT ATC
Ser Glu Val Asp Thr Gln Ser Thr Glu Asp Ser Glu Leu Glu Thr Ile
685 690 695

TCA GTC ACT GCA GAA AAA GTT AGA GAT CGT AAA GAT AAT GAA GTA ACT
Ser Val Thr Ala Glu Lys Val Arg Asp Arg Lys Asp Asn Glu Val Thr
700 705 710

GGA CTT GGC AAA ATT ATA AAA ACT AGT GAA AGT ATC AGC CGA GAA CAA
Gly Leu Gly Lys Ile Ile Lys Thr Ser Glu Ser Ile Ser Arg Glu Gln
715 720 725

GTA TTA AAT ATT CGT GAT CTA ACA CGC TAT GAT CCA GGG ATT TCA GTT
Val Leu Asn Ile Arg Asp Leu Thr Arg Tyr Asp Pro Gly Ile Ser Val
730 735 740

GTA GAA CAA GGT CGC GGT GCA AGT TCT GGA TAT TCT ATT CGT GGT ATG
Val Glu Gln Gly Arg Gly Ala Ser Ser Gly Tyr Ser Ile Arg Gly Met
745 750 755 760

FIG.3F.

```
TTC TTA TTA GGT CTC CGT ACT CCC AGT TCT GAA ATA CCT AAA GAA GGA
Phe Leu Leu Gly Leu Arg Thr Pro Ser Ser Glu Ile Pro Lys Glu Gly
  480                 485                 490

AGT GCA AAA TAT CAT GGT AAT TGG TTT GGT TAT ATT AGT GAT GGC GAG
Ser Ala Lys Tyr His Gly Asn Trp Phe Gly Tyr Ile Ser Asp Gly Glu
  495                 500                 505

ACA TCT TAC TCC GCC AGT GGT GAT AAG GAA CGC AGT AAA AAT GCT GTC
Thr Ser Tyr Ser Ala Ser Gly Asp Lys Glu Arg Ser Lys Asn Ala Val
510                 515                 520                 525

GCC GAG TTT AAT GTA AAT TTT GCC GAG AAA ACA TTA ACA GGC GAA TTA
Ala Glu Phe Asn Val Asn Phe Ala Glu Lys Thr Leu Thr Gly Glu Leu
                530                 535                 540

AAA CGA CAC GAT ACT CAA AAT CCC GTA TTT AAA ATT AAT GCA ACC TTT
Lys Arg His Asp Thr Gln Asn Pro Val Phe Lys Ile Asn Ala Thr Phe
  545                 550                 555

CAA AGT GGT AAG AAT GAC TTC ACT GGT ACA GCA ACC GCA AAA GAT TTA
Gln Ser Gly Lys Asn Asp Phe Thr Gly Thr Ala Thr Ala Lys Asp Leu
  560                 565                 570
```

FIG.3D.

```
CAT CCC TTT ACC AGC GAG GGA ACA TTA GAA GGT GGT TTT TAC GGG CCT
His Pro Phe Thr Ser Glu Gly Thr Leu Glu Gly Gly Phe Tyr Gly Pro
                290                 295                 300

GAG GGT CAA GAA TTA GGA GGA AAG TTT TTA GCT CAC GAC AAA AAA GTT
Glu Gly Gln Glu Leu Gly Gly Lys Phe Leu Ala His Asp Lys Lys Val
            305                 310                 315

TTG GGG GTA TTT AGT GCC AAA GAA CAG CAA GAA ACG TCA GAA AAC AAA
Leu Gly Val Phe Ser Ala Lys Glu Gln Gln Glu Thr Ser Glu Asn Lys
        320                 325                 330

AAA TTA CCC AAA GAA ACC TTA ATT GAT GGC AAG CTA ACT ACT TTT AAA
Lys Leu Pro Lys Glu Thr Leu Ile Asp Gly Lys Leu Thr Thr Phe Lys
    335                 340                 345

ACA ACC AAT GCA ACA GCC AAT GCA ACA ACC GAT GCA ACA ACC AGT ACA
Thr Thr Asn Ala Thr Ala Asn Ala Thr Thr Asp Ala Thr Thr Ser Thr
350                 355                 360                 365

ACA GCC AGT ACA AAA ACC GAT ACA ACA ACC AAT GCA ACA GCC AAT ACA
Thr Ala Ser Thr Lys Thr Asp Thr Thr Thr Asn Ala Thr Ala Asn Thr
                370                 375                 380
```

FIG.3B.

```
ATT GAA GAG GAT GTT AAA AAT GAC AAT CAA AAC GGC GAG CAC CCT ATT
Ile Glu Glu Asp Val Lys Asn Asp Asn Gln Asn Gly Glu His Pro Ile
 95                  100                 105

GAC TCA ATA GTC GAT CCT AGA GCA CCA AAT TCA AAC GAA AAT CGT CAT
Asp Ser Ile Val Asp Pro Arg Ala Pro Asn Ser Asn Glu Asn Arg His
110                 115                 120                 125

GGA CAA AAA TAT GTA TAT TCA GGG CTT TAT TAT ATT CAA TCG TGG AGT
Gly Gln Lys Tyr Val Tyr Ser Gly Leu Tyr Tyr Ile Gln Ser Trp Ser
                130                 135                 140

CTA AGA GAT TTA CCA AAT AAA AAG TTT TAT TCA GGT TAC TAT GGA TAT
Leu Arg Asp Leu Pro Asn Lys Lys Phe Tyr Ser Gly Tyr Tyr Gly Tyr
            145                 150                 155

GCG TAT TAC TTT GGC AAT ACA ACT GCC TCT GCA TTA CCT GTA GGT GGC
Ala Tyr Tyr Phe Gly Asn Thr Thr Ala Ser Ala Leu Pro Val Gly Gly
        160                 165                 170

GTA GCA ACG TAT AAA GGA ACT TGG AGC TTC ATC ACC GCA GCT GAA AAT
Val Ala Thr Tyr Lys Gly Thr Trp Ser Phe Ile Thr Ala Ala Glu Asn
    175                 180                 185
```